UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

FELIPE DE SOUZA OLIVEIRA

Production Of Green Hydrogen in Brazil: A Financial and Temporal Analysis

Juiz de Fora
2025

FELIPE DE SOUZA OLIVEIRA

Production Of Green Hydrogen in Brazil: A Financial and Temporal Analysis

Dissertação apresentada ao Programa de Pós Graduação em Economia Aplicada da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Economia Aplicada.

Orientador: Rafael Morais de Souza
Coorientador: Victor Eduardo de Mello Valério

Juiz de Fora
2025

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

De Souza Oliveira, Felipe .
Production Of Green Hydrogen in Brazil : A Financial and Temporal Analysis / Felipe De Souza Oliveira. -- 2025.
58 p. : il.

Orientador: Rafael Morais de Souza
Coorientador: Victor Eduardo de Mello Valério
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Economia. Programa de Pós-Graduação em Economia, 2025.

1. Energia Renováveis. 2. Hidrogênio Verde. 3. Análise de Investimentos. 4. Séries Temporais. 5. Previsão. I. Morais de Souza, Rafael , orient. II. Eduardo de Mello Valério, Victor, coorient. III. Título.

Felipe de Souza Oliveira

Production Of Green Hydrogen in Brazil: A Financial and Temporal Analysis

Dissertação apresentada ao Programa de Pós-graduação em Economia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Economia Aplicada. Área de concentração: Economia.

Aprovada em 20 de fevereiro de 2025.

BANCA EXAMINADORA

Dr. Rafael Morais de Souza - Orientador

Universidade Federal de Juiz de Fora

Dr. Victor Eduardo de Mello Valério - Coorientador

Universidade Federal de Itajubá

Dr. Douglas Sad Silveira

Universidade Federal de Juiz de Fora

Dr. Pedro Paulo Balestrassi

Universidade Federal de Itajubá

Juiz de Fora, 12/02/2025.

Documento assinado eletronicamente por **Rafael Morais de Souza**, **Professor(a)**, em 20/02/2025, às 11:56, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Victor Eduardo de Mello Valerio**, **Usuário Externo**, em 20/02/2025, às 12:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Douglas Sad Silveira**, **Professor(a)**, em 20/02/2025, às 14:52, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **PEDRO PAULO BALESTRASSI**, **Usuário Externo**, em 20/02/2025, às 16:40, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2242320** e o código CRC **2A464114**.

AGRADECIMENTOS

A conclusão desta dissertação representa não apenas a realização de um objetivo pessoal, mas também o reflexo do apoio e dedicação de muitas pessoas que estiveram ao meu lado durante esta jornada.

Em primeiro lugar, agradeço à minha família, que foi meu maior suporte em todos os momentos. A vocês, que sempre acreditaram em mim, ofereço minha eterna gratidão pelo amor incondicional, pela paciência nos momentos difíceis e pelo incentivo constante, que foram fundamentais para que eu não desistisse diante dos desafios.

Aos meus professores e orientadores, expresso minha profunda admiração e respeito. Obrigado por compartilharem comigo o conhecimento, pela orientação criteriosa e pelo comprometimento com meu desenvolvimento acadêmico. Suas contribuições foram essenciais para a construção deste trabalho, e levo comigo os ensinamentos que me transmitiram, tanto no âmbito profissional quanto pessoal.

À Universidade Federal de Juiz de Fora, sou grato por me proporcionar uma formação de qualidade e um ambiente acadêmico desafiador e inspirador. A infraestrutura, os recursos disponibilizados e as oportunidades de aprendizado foram determinantes para o avanço da minha pesquisa e para a minha trajetória acadêmica.

A Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) que apoiou o desenvolvimento do presente trabalho por meio do financiamento do projeto intitulado “Impactos Econômicos da Geração Solar Fotovoltaica Distribuída no Estado de Minas Gerais: Uma Abordagem de Modelagem Inter-Regional de Equilíbrio Geral Computável (RED-00161-22)” aprovado no edital Edital Nº 007/2021 - Redes de Pesquisa Científica e Desenvolvimento Tecnológico com Foco em Demandas Estratégicas, estruturado pela FAPEMIG e executado pela Rede de Análise em Energia Renovável e Desenvolvimento Econômico – RAERDE.

Por fim, agradeço a todos aqueles que, direta ou indiretamente, contribuíram para a realização deste trabalho. A cada gesto de apoio, palavra de incentivo e troca de experiências, deixo meu mais sincero obrigado.

A todos vocês, minha eterna gratidão e reconhecimento.

RESUMO

Esta dissertação aborda a produção de Hidrogênio Verde (H_2) no Brasil. O primeiro objetivo é calcular a viabilidade financeira de uma planta de H_2 Verde utilizando a métrica do Custo Nivelado de Hidrogênio (LCOH) para municípios brasileiros. O propósito é analisar como o preço varia ao longo do território nacional. Em seguida, busca-se desenvolver um modelo híbrido de previsão horária para a irradiação solar, considerando o município com o menor e maior LCOH. Ao combinar esses dois métodos, esta dissertação visa ampliar a compreensão sobre a produção de H_2 Verde no Brasil, fornecendo, assim, informações relevantes para o desenvolvimento do setor.

Palavras-chave: Energia Renováveis, Hidrogênio Verde, Análise de Investimentos, Séries Temporais.

ABSTRACT

This dissertation addresses Green Hydrogen (H_2) production in Brazil. The first aim is to calculate the financial feasibility of a Green H_2 plant using the Levelized Cost of Hydrogen (LCOH) metric for Brazilian municipalities. The objective is to analyze how the price varies throughout the national territory. After that, the aim is to develop a hybrid hourly forecasting model for solar irradiation, considering the municipality which have the lowest and the highest production cost. By combining these two methods, this dissertation seeks to enhance the understanding of Green H_2 production in Brazil, thus providing relevant information for the development of the sector.

Keywords: Renewable Energy, Green Hydrogen, Investment Analysis, Times series.

List Of Figures

List Of Tables

Table of contentes

1 INTRODUCTION

Human-caused climate change is already impacting every region of the world in various ways and has become a critical issue in contemporary society. According to the latest Intergovernmental Panel on Climate Change (IPCC, 2023) report, global surface temperatures have risen by 1.1°C over the past decade. This alarming trend highlights the urgent need to improve the relationship between greenhouse gas emissions and environmental sustainability. Among the various sectors of the modern economy, energy and electricity generation is regarded as one of the most important (IPCC, 2014), however, Ritchie et al. (2020) and Asmelash et al. (2020) states that it is also one of the primary sources of greenhouse gas emissions on a global scale, along with heat production and also followed by transport, manufacturing and construction. Through the decades a wide range of energy sources have been utilized, and as presented by Science Direct (2016), the importance of an energy source is determined by the product of the energy available and its efficiency for organic compound synthesis. To exemplify, in 2024, 60.65% of global electricity generation was derived from fossil fuels, while 30% was derived from renewable sources, with hydropower, wind, and solar ranked in descending order of importance (Ember, 2024).

Brazil represents a notable exception, with a predominantly clean electricity generation profile. As reported by the National System Operator (ONS, 2024), 88.6% of Brazil's electricity matrix is composed of renewable sources, with 48.4% generated by hydropower, 14.1% by wind, and 9.1% by solar. The high dependency on hydropower has recently raised concerns due to the uncertain rainfall patterns, which result in significant variability in water levels within the country's primary reservoirs, as Chala et al. (2019) discusses and may lead to challenges in maintaining the supply capacity electric (Tavares, 2023). For better comprehension, hydroeletric power, also kown as hydropower, is defined by Killingtveit (2020) as renewable energy source where electrical power is derived from the energy of water moving from higher to lower elevations. Actions were carried out with the aim of increasing electricity production from alternative sources, such as the Incentive Program for Alternative Sources of Electric Energy (PROINFA) established by Law No. 10,438/2002, which primarily focused on wind energy generation (Brazil, 2025). Additionally, regulatory resolutions N482/2012 and N687/2016, structured by the National Electric Energy Agency (ANEEL) (ANEEL, 2024), along with Federal Law No. 14,300/2022, represent the legal framework that fostered the development of photovoltaic energy in Brazil (Brazil, 2025).

Photovoltaic energy refers to an electricity generation system based on the direct conversion of solar radiation into electricity using photovoltaic cells. These cells are composed of semiconductor materials, usually silicon, that generate an electric current when exposed to sunlight through the photovoltaic effect (Omer, 2008). Solar energy has experienced exponential growth over the last decade, with an increase in installed capacity from 23 MW in 2014 to approximately 53,191MW in 2024 (ABSOLAR,2024). As posited by Lopes et al. (2023), Brazil is endowed with considerable potential for solar power generation largely due to its tropical location, expansive territory, and elevated levels of direct solar irradiation. Furthermore, it is worth noting that, in addition to a higher average daily irradiation (5.5 kWh/m² in the Northeast Region) if compared to European countries such as Spain (4.95 kWh/m²), France (3.7 kWh/m²), and Germany (3.0 kWh/m²), Brazil also experiences less variation throughout the year. This results in more stable energy generation

over the year (Pereira et al., 2017). Although renewable sources tend to be initially more expensive, as they expand, there are technological and scale gains that lead to cost reductions, resulting in a reduction in the cost per unit of energy produced. Consequently, it is projected that by 2030, the CAPEX of a photovoltaic system will be around 300 USD/kW, taking as examples China and Germany, two key countries in this sector, with this value expected to decrease to 250 USD/kW by 2050.

However, the integration of renewable energy sources into the power grid brings additional challenges due to their intermittent and seasonal nature, which results in unpredictable variations in energy generation. This characteristic can lead to imbalances between supply and demand, affecting the stability and reliability of the electrical system, particularly in scenarios with high penetration of these technologies (IEA, 2023). In this context, green hydrogen emerges as a strategic alternative to address these limitations. Green or low-carbon hydrogen (H_2) refers to hydrogen produced from renewable energy sources, without greenhouse gas emissions during the process. The most common route for its production is water electrolysis, where the electricity required to split water molecules into hydrogen (H_2) and oxygen (O_2) comes from clean sources, such as solar, wind, or hydropower (Oliveira et al., 2021). Green H_2 not only reduces greenhouse gas emissions but also plays a crucial role in storing surplus energy generated by intermittent sources. During periods of high wind or solar radiation availability, excess electricity can be used to produce hydrogen, which can then be stored for extended periods in pressurized tanks, salt caverns, or other advanced chemical storage methods (IRENA, 2022). Moreover, green hydrogen can be reconverted into electricity using fuel cells or modified gas turbines, serving as a flexibility solution for the power grid. This storage and reconversion capability not only provides a way to address intermittency but also acts as a tool for managing energy demand, reducing the reliance on fast-response fossil fuel generation, such as natural gas (Maestre et al., 2021).

Therefore, it can be observed that the use of photovoltaic energy is closely related to the production of green H_2. This has gained attention in recent years as a mean to support not only the energy transition but also the decarbonization of various economic sectors. Brazil is positioned as one of the main players in the green H_2 economy, with the first production plants already under construction (Farrel, 2023). Additionally, a plan has already been developed by the Ministry of Mines and Energy (MME), outlining goals and objectives for the sector in the coming years (MME, 2023).

However, its large-scale implementation faces economic and technological challenges, such as high initial production costs (CAPEX) and the limited efficiency of electrolysis systems. One of the main factors affecting production costs, considering green hydrogen production through photovoltaic energy, is the availability of solar resources. This resource varies significantly not only throughout the national territory, but also between days - due to the seasons - and between hours of the day - due to the position of the sun. (Pereira *et al*,2017). As a result, the cost of H_2 is expected to vary across the country, since, for example, locations with lower average solar irradiation require larger photovoltaic plants, increasing CAPEX and, consequently, the cost of hydrogen, and vice versa.

One challenge associated with the generation of electricity from renewable sources is the inherent inconsistency in output over time, since such generation depends on stochastic factors, such as solar irradiation. This variability, when not analyzed properly, generates the need to oversize photovoltaic systems and storage systems in green H_2 production plants,

increasing significantly their production cost and, therefore, the financial viability of these projects.

Forecasting techniques help mitigate this issue by providing insights not only about variations between hours of the day but also variations between days in the electricity production. In the case of the literature on renewable energy sources, the are two principal classifications of. The first classification is based on the type of forecast, which can be direct or indirect. In direct forecasting, the historical production series is employed to generate predictions. In contrast, indirect forecasting utilizes the historical solar irradiation series to estimate production, taking the solar plant's size into account (Yang et al, 2021).

The second classification is based on temporal considerations. Ultra-short-term forecasting predicts values for the subsequent few hours, short-term forecasting predicts values within a day, and long-term forecasting covers periods exceeding a week, potentially up to a year (Liu et al., 2022). The utility of each type of forecast is distinct. For example, long-term forecasting provides insight about the potential generation capacity of a location, which informs decisions regarding the feasibility of implementing a new solar plant (Yang et al., 2021).

Another aspect of modeling in this area is the complexity of the data, which has led to the use of hybrid models (Wang et al., 2021; La Tona & Di Piazza, 2023; Rosen et al., 2023). These models integrate two or more data processing techniques to reduce the occurrence of prediction errors (Xiao et al., 2023). For instance, statistical techniques that differentiate between noise and signal in time series, such as Wavelet Transformation or Singular Spectrum Analysis (SSA), can be integrated with Artificial Neural Networks (ANN) (Jian & Agelidis, 2015; Wang et al., 2021; Zhu et al., 2022).

In this context, this dissertation aims to develop an indirect, short-term, hybrid forecasting model for hourly solar irradiation forecasting. To this end, the dissertation was organized into the following specific objectives:

- Survey of equilibrium prices for the production of green hydrogen through electrolysis using a photovoltaic system in each Brazilian municipality;
- Selection of municipalities to implement the forecast models based on financial viability criteria;
- Survey of irradiation time series in selected municipalities;
- Implementation of the hybrid forecasting model (SSA+ decision tree);
- Implementation of the hybrid forecast model (SSA+ ANN);
- Comparison between forecasting models;
- Selection of the best performing model.

To identify the most suitable locations, a filtering process is applied, analyzing the financial viability for green hydrogen production Brazilian’s municipalities, using the Levelized Cost of Hydrogen (LCOH) as a key metric of equilibrium price. To measure LCOH conventional calculation techniques will be used, without the use of a forecast model.

The contribution of this study lies in expanding knowledge on photovoltaic generation in short-term forecasting and improving the understanding of the feasibility of green hydrogen production in Brazil. It is noteworthy that the existing literature does not comprehensively

address this issue in the Brazilian context, highlighting a gap to be filled. The results of this study can benefit different sectors of society. Consumers may benefit from the potential reduction in green hydrogen production costs, making this energy source more accessible and contributing to lower electricity bills in the medium and long term. Energy companies, in turn, can use more accurate forecasting models to optimize the operation of their photovoltaic plants, reducing waste and improving efficiency in power generation and distribution. The government can also benefit from more detailed data to support public policies that promote renewable energy and the hydrogen economy, enhancing energy security and reducing reliance on fossil fuels. From a macroeconomic perspective, greater predictability in renewable energy generation can attract investments to the sector, boosting Brazil's competitiveness in the global green hydrogen market and contributing to sustainable economic growth.

Additionally, this study is justified by the fact that regional differences in solar resource availability affect the production cost of green hydrogen, thereby impacting its price competitiveness across Brazilian municipalities. Furthermore, given its temporal variability, forecasting models can be valuable tools for understanding these variations. Consequently, they tend to mitigate uncertainties and support the development of the green hydrogen economy in Brazil.

In addition to the introductory section, this work is organized as follows: Chapter 2 presents a bibliographic review; Chapter 3 explains the methodology used; in Chapter 4, the Results are demonstrated; finally, Chapter 5 concludes this dissertation;

2 BIBLIOGRAPHIC REVIEW

2.1 Bibliographic Review of Forecast

Forecasting photovoltaic energy is a challenging task due to the stochastic and intermittent generation of energy caused by local climatic conditions such as sun exposure, weather, and season (Liu et al., 2022). Forecasting aims to provide information for the maintenance of photovoltaic plants, helping energy companies and consumers better understand the peculiarities of energy generation. This includes the amount of energy generated, its variation over time, and how energy production meets demand throughout the day.

Photovoltaic generation can be forecasted in two ways. Direct forecasting involves using the plant's own historical series to generate the forecast, while indirect forecasting entails forecasting the irradiation level at the plant's location before converting it into generation, given the installed capacity of photovoltaic generation (Jian & Agelidis, 2015).

Forecasting can be classified based on the time horizon. According to the literature, there are three types of forecasts: very short-term, short-term, and long-term. A very short-term forecast has a time horizon of no more than one day, and due to the short time horizon, it experiences the largest fluctuations in production. This type of forecast helps companies in the sector to reduce the problem of intermittent generation and grid connection issues (Jian & Agelidis, 2015). The short-term forecast is one in which the forecast horizon is generally one day ahead. Its usefulness lies in daily generation planning and system maintenance (La Tona & Di Piazza, 2023). Finally, the long-term forecast is one in which the forecast horizon is longer than 72 hours, ranging from months to a year. Its utility lies in choosing a new location for a plant and its economic viability (Jian & Agelidis, 2015).

First, Parmezan et al. (2019) discusses the state of the art regarding the main models for forecasting time series. In this regard, the main components of a time series (trend, seasonality, and error component), the concept of stationarity, and how each of them affects forecasting are presented, followed by an analysis of parametric forecasting models. According to the authors, these types of models require a priori knowledge of the distribution of the data and from this a set of parameters is estimated to make the forecast, among which we can highlight the SARIMA and Exponential Damping models. The second class of methods described by the authors are the so-called nonparametric methods, which describe the data without any a priori knowledge of its distribution. Furthermore, they do not explicitly depend on the estimation of parameters to describe the behavior of the series. Examples of methods in this category are Artificial Neural Networks and Support Vector Machines (SVM).

Liu et al (2022) review forecasting methods for the photovoltaic sector, both in terms of time and space, i.e. how geographic space affects forecasting and how photovoltaic plants in the same location can take advantage of this fact to generate more robust forecasts. The authors begin by describing the classification of forecasts in the sector, both in terms of form

(direct or indirect) and time horizon (very short, short and long term). In addition, the authors explain the intention and possible difficulties of each of them, for example, in the very short term forecast we have that the forecast horizon varies between zero and four hours and that the main difficulty is that the stochastic characteristics and variability of photovoltaic generation occurs more markedly in this period, which generates a strong noise component that makes the forecast difficult, this fact can be explained by the appearance, movement and dissipation of clouds at the generation site, which generates a considerable fluctuation in generation. To mitigate this, the authors suggest using images of the generation site as one of the variables in the model, so that the model can identify how the clouds affect the forecast. In addition, considering the forecast for several generation sites in the same geographical space, the use of clustering methods for these images is recommended to improve the quality of the forecast, the example given by the authors is the K-neighbors method as a form of clustering.

Another review of photovoltaic forecasting methods is given in Carneiro et al. (2022), where the authors classify forecasting models into four possible approaches. The first two are the statistical or parametric approach and the nonparametric approach; these two classifications are similar to those in the article by Parmezan et al. (2019). The third approach is what the authors call the physical approach, where the goal of the model is to describe the physical relationship between the variables. For example, the relative humidity and wind speed are calculated and used as input variables for the photovoltaic prediction model. The last type of approach advocated by the authors are the so-called hybrid models, which consist of using more than one statistical technique for forecasting, such as the use of clustering techniques like "k-neighbors" to group data into similar groups and then use these groups in forecasting models, such as an artificial neural network. Another example of hybrid models is the use of time series decomposition techniques such as "Singular Spectrum Analysis (SSA)" or "Wavelet Transformation" where the original series is decomposed into trend, seasonality and error components and then each of these series is used in forecasting models. According to the authors, the latter approach has grown in recent years due to its ability to produce more accurate forecasts.

Another meta-analysis of forecasting models in the photovoltaic sector is carried out by Radzi et al (2023). In this paper, the concepts of forecast horizon, forecast classification, and the possible approaches of the previously discussed models are discussed. In this sense, we can highlight the architecture of Recurrent Neural Networks (RNN), which, according to the authors, this type is relevant in the context of time series because there is the possibility that the prior knowledge of one neuron is used by the others, and since we are working with sequence data, this is a desirable characteristic. The authors report that among the various RNN architectures, one of the most widely used in the literature is called Long-Short Term Memory (LSTM). The authors also point out that among the studies analyzed, the most common variables used in forecasting are: the photovoltaic generation itself, when we are working with direct forecasting. When we are working with indirect forecasting, some measure of solar irradiation is used, for example, global solar irradiation linked to meteorological variables such as relative humidity, ambient temperature, wind speed, atmospheric pressure, etc.

Jian and Agelidis (2015) use a hybrid model to forecast solar PV generation in the short term, using data on solar PV generation itself and meteorological variables such as solar irradiation intensity and maximum and minimum temperatures. The techniques used by the authors to build the model can be divided into three parts. First, the Wavalet transformation was used to extract the high and low frequency coefficients. This data is then fed into an

artificial neural network model. Finally, Markov chains are used to correct the errors. In this way, according to the authors, the proposed model is able to make good predictions, with a MAE (Mean Absolute Error) of 0.015.

In Zhaoxuan Li et al. (2016), the authors compare two recurring models in the photovoltaic forecasting literature: Artificial Neural Networks and Support Vector Machines. In addition, a simplified approach is developed using reduced exogenous inputs without solar irradiance to forecast photovoltaic production 15 minutes, 1 hour, and 24 hours in advance. The forecasting model, developed by ANN and Support Vector Machine (SVM), respectively, predicts the power output based on the historical record of PV and online weather services.

Succetti et al (2020) explain how the Long-Short Term Menory (LSTM) architecture is used for forecasting in the photovoltaic sector. According to the authors, it has the characteristic of dealing with the long-term temporal dependence present in photovoltaic series, such as the ability of these models to learn how different types of weather affect photovoltaic generation. In this sense, the authors propose and analyze four variations of the LSTM architecture: "C-LSTM", "Conv-LSTM", "Mult-LSTM" and "Stacked-LSTM". In this experiment, the forecast horizon is considered to be one hour ahead, and the variables used to build the model are: solar irradiance, air temperature, and wind speed. In conclusion, the authors found that multivariate models have a lower Mae than univariate models. However, the authors also point out that it is necessary to consider not only the performance metrics, but also the complexity of the model and the time required to train it in order to select the most appropriate model.

Yang et al (2021) analyze the complementarity of photovoltaic and wind energy sources in a system. Complementarity has become a way to reduce the variability of renewable energy generation and thus make generation systems more stable over time. In this article, the authors use the Singular Spectrum Analysis (SSA) technique to decompose the series into their respective components (oscillatory, trend, and noise). A complementarity index is then computed. The authors conclude that there is a correlation between the series.

In Wang et al (2021), the authors analyze how different climates affect the accuracy of forecasting models. Days were classified as: sunny, cloudy, slightly rainy, and very rainy. Similar days were then grouped together and used to train different models for each day. According to the authors, solar irradiation and atmospheric temperature are the main factors affecting photovoltaic energy production. It is worth noting that the authors also used the "SSA" technique to decompose the series into trend, oscillation and error in order to improve the accuracy of the prediction. The variables used in the model were meteorological data from the day of the forecast and historical data from similar days, as well as photovoltaic generation data. The forecast horizon used was 15 minutes.

Jin and Zhang (2021) propose a short-term forecasting model for photovoltaic generation using photovoltaic production and meteorological factors as variables. The authors use two methods to minimize the noise component after decomposing the series using the SSA technique: "Fuzzy Information Granulation" and "Improved Cuckoo Algorithm (ICS)". The forecasting model used was a neural network called "Extreme Learning Machine", which consists of a "Feedforward" architecture with a hidden layer. The authors concluded that the inclusion of meteorological variables in the model improves its predictive capacity and that it has the best predictive capacity on sunny days, while it has the largest error on rainy days.

In Oliveira (2022), one of the author's objectives is to study the new load profiles of Brazilian consumers, mainly due to the growth of micro and mini distributed photovoltaic generation. To this end, she proposes her own methodology for estimating photovoltaic generation in each of the subsystems that make up the National Integrated System (SIN), the process comprises four stages. Initially, representative municipalities are selected on the basis of distributed generation (DG) installations for each subsystem. Subsequently, meteorological data is collected for these municipalities. Thereafter, the operation of GD systems is simulated with the assistance of meteorological bases. Finally, the results are weighted according to the weight of each municipality in the subsystem. It also discusses the peculiarities of energy consumption on atypical days, such as holidays, and how these pose a challenge in terms of forecasting. Finally, it is explained that there is more than one seasonal component in the series, in this case there is daily and weekly seasonality, and a model is proposed that addresses this issue, the Double Seasonal Holt-Winters Model.

In Zhu et al. (2022), the authors analyze several "traditional" energy price series, such as coal, compared to polysilicon, an important component in the production of photovoltaic panels. According to the authors, the SSA technique is used to improve the accuracy of the models by improving the prediction of the trend and error components of each series. After decomposition, the series are used to feed a long-term memory artificial neural network model. They also discuss how the variables relate to an indicator of carbon emission intensity.

Xiao et al (2023) discuss how hybrid models, i.e. those that combine more than two or more techniques, are used in the field of photovoltaic energy forecasting. In addition, a comparison is made between different Artificial Neural Network (ANN) architectures with the aim of selecting the best one for forecasting in the one hour and one day ahead horizons. The variables used by the authors were: photovoltaic energy production, temperature, relative humidity, irradiation and precipitation. According to the authors, the combination of Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM) architectures makes it possible to extract the non-linear part of the series, increasing the accuracy of the model. In addition, a comparison of the performance of the proposed model in different seasons is made.

In Rosen et al. (2023), the authors use an image capture method to create a grouping, "clusters", which in turn serve as one of the variables for the model, in addition to photovoltaic generation and meteorological variables such as: solar irradiation, temperature, relative humidity, etc. Markov chains are used to improve the flexibility of the model. In addition, Markov chains are used to improve the flexibility of the model. In this sense, this technique works with a transition matrix, which means that, given a certain parameter, there is a transition from one chain to another; it is also worth noting that this approach uses Bayesian statistics, so that probabilistic forecasts are generated based on their respective probability distributions.

2.2 Photovoltaic Sector in Brazil

In 2024, the installed capacity in Brazil reached 200 gigawatts (GW). The generation of electrical energy in Brazil is predominantly derived from clean and renewable sources, with the majority of this energy generated from hydroelectric sources, which accounted for 49.2% of the total energy generated in that year (ONS, 2024). Nevertheless, in recent years, there has been a debate about the excessive dependence of the Brazilian electric matrix on

hydroelectric power, which could jeopardize the country's energy security, given an increasingly uncertain rainfall regime (Tavares, 2023). In accordance with the necessity for diversification of the national electric matrix, there has been an increase in the generation of other clean and renewable sources of energy. Firstly, the growth of wind generation in recent years, particularly in the Southern and Northeastern regions, has led to an increase in the importance of this source of energy in meeting market demand (ONS, 2024). In aggregate terms, wind generation accounts for 13% of the Brazilian electric matrix (ABSOLAR, 2024).

Another noteworthy source of energy generation is photovoltaic. this source has experienced exponential growth over the past decade, increasing from 13 MW of installed capacity in 2013 to approximately 38 MW in 2023. As stated by Lopes et al. (2023), Brazil is endowed with considerable potential for photovoltaic electrical generation, due to its tropical location, the size of its territory, and the level of direct solar irradiation received. Moreover, while renewable sources may initially be more expensive, as they expand, there are scale and technological gains, reflected in the drop in prices of these equipment. In this context, the Financial Times reports that the price of solar panels reached its lowest point in 2024, with the price dropping from 0.8 dollars per watt to 0.2 dollars per watt over the past decade.

In Brazil, photovoltaic energy generation can be classified in two distinct ways. The first category is centralized generation, which involves the construction of large-scale generating plants in locations with a high level of solar irradiation, typically outside urban areas. The objective of this approach is to meet the general energy demand (Lopes et al., 2023). As illustrated in Figure 1, centralized generation accounts for 31% of the installed photovoltaic potential in 2024. Conversely, distributed generation can be defined as electrical generation carried out in close proximity to or in conjunction with the consumer, and it is directly connected to the energy distribution networks through installations at consumer units (Rodrigues et al., 2020).

The advent of distributed generation (DG) in Brazil is marked by Regulatory Resolution No. 482/2012 of ANEEL in 2012. This resolution establishes the possibility for consumers to generate their own energy through renewable sources and, if necessary, to supply the surplus to the distribution network (ANEEL, 2024). Moreover, the resolution delineates the criteria for categorizing photovoltaic generation facilities into two distinct categories: microgeneration, which applies when the installed capacity is equal to or less than 75 kW, and minigeneration, which applies when the installed capacity is between 75 kW and 5 MW (Cardoso, 2021). Moreover, the Net Metering System was established, enabling consumers to install small generators, such as solar panels or wind turbines, in their consumer units and exchange energy with the local distributors with the intention of offsetting the value of the energy generated by these units from their electricity bill. In this context, these credits are valid for up to 60 days and may be utilized in various consumer units, provided that they are owned by the same entity (Villanueva, 2023).

In 2015, the National Electric Energy Agency (ANEEL) implemented changes to sector regulations through Normative Resolution No. 687/2015. In this resolution, the maximum installed capacity for an enterprise to be considered a mini-generation facility was reduced to up to 75 kW, while micro-generation classification was set between 75 kW and 5 MW. Furthermore, microgeneration enterprises are now permitted to include individual properties, condominiums, and cooperatives. Consequently, a novel consumption modality was introduced, designated as "remote self-consumption." This modality occurs when the generated energy exceeds the consumed energy, allowing the consumer to have available

credits. Such credits may be utilized over a period of 30 to 60 months. Consequently, these credits may be used to offset consumption in another unit, provided that the fllowing conditions are met: firstly, the other consumer unit must be under the same CPF[1] or CNPJ[2]; and secondly, both units must be within the same concession area of the distributor (ANEEL, 2024).

In this context, the concept of "shared consumption" emerges. This is when consumers within the same concession area form an association through a consortium, in the case of individuals (PF[3]), or a cooperative, in the case of legal entities (PJ[4]), and use credits from the surplus generated by micro or mini photovoltaic generation projects in different locations from where the energy was originally generated. This modality has created new business opportunities in the sector, including energy purchase agreements, leasing of enterprises, and solar condominiums (Luna et al., 2019).

Due to the rapid development of the sector in recent years, it has become necessary to regulate it through legislation rather than a Normative Resolution from ANEEL. This led to the enactment of Law No. 14,120 on March 1, 2021. The most significant change introduced by this law, compared to previous Normative Resolutions, is in the application of discounts on the electricity tariff for photovoltaic projects. Specifically, it modifies the discount related to the Wire Tariff (TUSD), one of the components of the electricity tariff. The TUSD reflects the costs for energy transportation and maintenance through both high and low-voltage cables. Importantly, this discount change only applies to projects that request a grant within twelve months from March 2, 2021, and that commence operations across all their generating units within 48 months from the grant date. In other words, projects that were operational before this deadline will continue to benefit from the tariff discount (Santos, 2022).

2.3 *Green Hydrogen*

The transition to a low-carbon economy has become a global priority as countries strive to reduce greenhouse gas (GHG) emissions and mitigate the impacts of climate change. In this context, green hydrogen has gained prominence as a clean and versatile energy carrier. Unlike conventional hydrogen production methods, green hydrogen is generated through water electrolysis using electricity from renewable sources such as solar and wind, ensuring a process free from direct carbon emissions (Sarker et al., 2023).

---

1 the CPF, or Individual Taxpayer Registry, is a unique identification number assigned to Brazilian citizens and resident foreigners. This number is issued by the Brazilian Federal Revenue Service (Receita Federal do Brazil) and is essential for a wide range of activities, such as opening a bank account, applying for a job, making major purchases, and filing taxes. The CPF consists of 11 digits and serves as a primary means of identifying individuals in Brazil's financial and administrative systems.

2 The CNPJ, or National Registry of Legal Entities, is a unique identification number for businesses and other legal entities operating in Brazil. Like the CPF, the CNPJ is issued by the Brazilian Federal Revenue Service. This number is crucial for businesses to conduct their activities legally, including opening bank accounts, issuing invoices, signing contracts, and paying taxes. The CNPJ consists of 14 digits and includes information about the business's legal structure, economic activity, and registration status.

3 "Pessoa Física" translates to "Natural Person" and refers to individual human beings. This category encompasses all citizens, including foreigners residing in Brazil, who are recognized as individuals with rights and obligations.

4 "Pessoa Jurídica" translates to "Legal Entity" and refers to organizations or entities that are legally recognized as having rights and responsibilities distinct from those of their members. Examples include corporations, partnerships, non-profit organizations, and government agencies.

One of the main drivers behind the increasing relevance of green hydrogen is the urgent need for decarbonization, which refers to the systematic reduction of carbon dioxide (CO_2) emissions in industrial processes, transportation, and energy generation. Fossil fuels, which currently dominate global energy consumption, are responsible for approximately 75% of total GHG emissions (IEA, 2023). Transitioning to low-carbon alternatives like green hydrogen is essential to achieving the targets set by the Paris Agreement, which aims to limit global warming to below 2°C above pre-industrial levels, with efforts to keep it within 1.5°C (Lee et al, 2023).

In parallel, energy security—the availability of reliable, affordable, and sustainable energy sources—has become a strategic priority for governments worldwide. The volatility of fossil fuel markets, geopolitical conflicts, and supply chain disruptions have highlighted the risks associated with dependence on imported hydrocarbons. Green hydrogen presents an opportunity to diversify the energy mix, enhance domestic energy independence, and reduce exposure to fuel price fluctuations. Countries with abundant renewable resources, such as Brazil, Australia, and Chile, have a strategic advantage in producing green hydrogen at scale, reducing reliance on fossil fuel imports while fostering economic growth through clean energy exports (Farrell, 2023).

Green hydrogen has a diverse range of applications across multiple sectors. In energy storage, it plays a crucial role in addressing the intermittency of renewable energy sources by storing excess electricity and releasing it when demand is high. In heavy industry, green hydrogen can be used to decarbonize steel production and green ammonia manufacturing, replacing fossil-fuel-based hydrogen. The transportation sector is another key area where hydrogen-powered fuel cells offer a viable alternative for heavy-duty vehicles, maritime shipping, and even aviation, which face challenges in direct electrification (Farrell, 2023).

Despite its potential, green hydrogen still faces significant economic and infrastructural challenges. The primary barrier is cost, as the electrolysis process remains considerably more expensive than conventional hydrogen production methods. Gray hydrogen, derived from natural gas through steam methane reforming (SMR), remains the dominant and more cost-competitive option. Other alternatives, such as turquoise hydrogen (produced via methane pyrolysis) and pink hydrogen (generated using nuclear energy), also offer lower production costs compared to green hydrogen. Additionally, infrastructure limitations—including production facilities, storage systems, and distribution networks—require substantial investment. Furthermore, concerns regarding storage and transportation safety, particularly due to hydrogen's low energy density and high flammability, necessitate technological advancements to ensure feasibility at scale (Mazzeo et al., 2022).

To enhance green hydrogen's competitiveness, countries have implemented strategic policies and investment frameworks to accelerate its adoption. The European Hydrogen Strategy aims to install at least 40 GW of electrolysis capacity by 2030, while Germany's National Hydrogen Strategy (Nationale Wasserstoffstrategie, NWS) is one of the most ambitious frameworks in Europe. Initially launched in 2020 and updated in 2023, outlines a structured plan to develop a competitive hydrogen market and position the country as a global leader in hydrogen technologies. The strategy allocates over €9 billion to hydrogen projects, with €7 billion dedicated to fostering domestic hydrogen production and €2 billion directed toward international partnerships, particularly in regions with high renewable energy potential, such as North Africa and Latin America. The plan emphasizes scaling up hydrogen

production, decarbonizing industries such as steel and chemicals, and expanding hydrogen infrastructure, including transport pipelines and refueling stations. By 2030, Germany aims to install at least 10 GW of domestic electrolysis capacity, supported by regulatory frameworks that facilitate demand creation through carbon pricing mechanisms and incentives for industrial adoption. Additionally, the H2Global initiative, backed by the German government, seeks to establish long-term purchase agreements for green hydrogen to secure supply chains and stabilize market prices. Germany is also integrating hydrogen into its energy system, leveraging offshore wind farms in the North Sea to power large-scale electrolysis plants (Roth *et al*, 2023).

China, the world's largest energy consumer and biggest emitter of CO_2, has also made substantial commitments to green hydrogen development. As part of its carbon neutrality goal by 2060, the Chinese government has integrated hydrogen into its Five-Year Plans, emphasizing its role in industrial decarbonization and clean transportation. The National Hydrogen Industry Development Plan (2021–2035) outlines China's ambition to build a domestic hydrogen supply chain, with an initial focus on reducing production costs and deploying hydrogen in transportation and heavy industry. Additionally, China leads in the manufacturing of low-cost electrolyzers, which are crucial for scaling up green hydrogen production. The country is expected to surpass 100 GW of installed electrolyzer capacity by 2035. In the transportation sector, China has committed to deploying 1 million hydrogen-powered vehicles by 2035, supported by government subsidies and state-backed industrial partnerships (Teng *et al, 2024).* Furthermore, provinces such as Inner Mongolia and Xinjiang are investing in large-scale projects that integrate wind and solar power with green hydrogen production, aiming to export hydrogen to neighboring Asian markets (Du *et al,* 2024).

In Brazil, the government has launched the National Hydrogen Program (PNH2) to foster research, technological development, and industrialization of green hydrogen. Several Brazilian states, including Ceará and Rio Grande do Norte, have signed agreements with international companies to develop hydrogen hubs, leveraging the country's abundant renewable energy resources. To attract investment, the government is considering tax exemptions, low-interest financing, and public-private partnerships, while the Brazilian Development Bank (BNDES) has introduced funding opportunities to further strengthen the sector with one of the cleanest energy matrices in the world, where over 80% of electricity comes from renewable sources, Brazil is uniquely positioned to become a global leader in green hydrogen production. Strategic investments in infrastructure, research and development, and regulatory frameworks will be essential to unlocking the country's full potential and fostering international partnerships in the hydrogen economy (De Andrade et al., 2024).

In summary, green hydrogen represents a pivotal opportunity for global decarbonization and long-term energy security. While economic and technological barriers remain, continued policy support, financial incentives, and advancements in electrolysis technology can position green hydrogen as a commercially viable solution. Given its exceptional renewable energy potential, Brazil has a unique opportunity to emerge as a key exporter of green hydrogen, strengthening its role in the global clean energy transition.

2.4 Levelized Cost of Hydrogen (LCOH)

As the purpose of this study is to derive an hourly hybrid forecasting model for solar irradiation to support H_2 production, it is necessary first to determine the most suitable

locations for its production, with production costs being the key factor in this selection. For this, the variable Levelized Cost of Hydrogen (LCOH) will be used, as expressed by Equation 1 (De Andrade et al., 2024). LCOH is a widely used metric for assessing the financial viability of H_2 projects, particularly those involving feasibility studies for plant construction (Dinh et al., 2023; Gallardo et al., 2021; Ramakrishnan et al., 2024).

$$\mathrm{LCOH} = \frac{\sum_{t=1}^{n} \frac{\mathrm{CAPEX}_t + \mathrm{OPEX}_t}{(1+r)^n}}{\sum_{t=1}^{n} \frac{G_{H_2}}{(1+r)^n}} \quad (1)$$

Where: CAPEX_t = Capital expenditures of the investment (US$); OPEX_t = Operational expenditures of the investment US$; n = system lifetime; r= discount rate (%); G_{H_2}= green hydrogen production (kg) in period n.

Intuitively, LCOH can be understood as the ratio between the sum of the present value of all costs over the plant's lifetime and the present value of hydrogen production. This concept is mathematically expressed by Equation 2.

$$LCOH = \frac{TOTEX_T}{\sum_{t=1}^{n} \frac{G_{H_2}}{(1+r)^n}} = \frac{Total\ Lifetime\ Cost}{Total\ Lifetime\ Hydrogen} \quad (2)$$

Where: $TOTEX_T$ = the total costs that result from the sum of CAPEX and OPEX (US$).

After calculating the LCOH, the municipalities will be selected, with the lowest and the highest production cost. With this, it becomes possible to obtain the time series of solar irradiation.

Using the selected time series, the next step is to decompose it using the Singular Spectrum Analysis (SSA) technique.

2.5 Singular Spectrum Analysis

The Singular Spectrum Analysis (SSA) technique is a method that has been employed in a multitude of disciplines, spanning from physics to the analysis of economic series (Santos, 2022). In the context of forecasting electricity generation from renewable sources, SSA has been widely employed to enhance the predictive capacity of models due to the inherent complexity of the data (Zhang et al., 2020; Jin and Zhang, 2021; Zhu et al., 2022). The objective of this methodology is to decompose a time series into a subset of signal series and then reassemble them, with the aim of extracting trend, oscillatory, and noise structure

components from the given series (Golyandina et al., 2018). One of the advantages of this method is that it does not require parametric restrictions, such as stationarity or normality. This makes it useful for non-stationary series, such as irradiation (Succetti et al., 2020). Therefore, it does not necessitate the requisite transformations to satisfy these assumptions, which can result in a considerable loss of information during the process. For instance, the potential loss of the trend component may occur when differentiation is applied (Hassani et al., 2013).

The SSA technique comprises a series of steps and decisions that must be made. These are the steps of decomposition and reconstruction, with the choices pertaining to the window length (L) and the number of eigenvalues (r). Furthermore, each of these steps is divided into two sub-steps. The embedding and singular value decomposition (SVD) are employed in the decomposition phase, while grouping and diagonal averaging are utilized in the reconstruction phase (Hassani et al., 2015).

Embedding:

Embedding is the initial procedure for time series analysis. Considering $z = [z_1, ..., z_n] \in \mathbb{R}$ as the series with N observations, embedding is performed by dividing the original series into subsets called observation windows (window length). Through the transformation $F: \mathbb{R}^N \rightarrow \mathbb{R}^{LxK}$ the following form is followed $X = [X_1, ..., X_k]^x_k \in \mathbb{R}^{LxK}$. In this case $X_j = [z_j, ..., z_{j+l+1}]^T \in \mathbb{R}^L$ with j = 1, 2, ..., K represents the number of vectors in the matrix X_j with K = N - L + 1, X and defined by the following structure (Korobeynikov et al., 2014):

In the case where F is an invertible map, the matrix X_j is referred to as the trajectory matrix, which is of Hankel-type[5], and the vectors X_j are known as lagged vectors. In this phase of the process, the sole requisite parameter is the parameter *L*.

$$X_j = F(z) = [X_1, \ldots, X_k]_k = (x_{ij})_{i=1,\, j=1} \quad (3)$$

The parameter *L* is related to the maximum number of rows that can be obtained in the constructed Hankel matrix. Therefore, it can be understood as the maximum number of eigentriples that can be extracted. It is of great importance for estimating components with low w-correlation between them. Selecting a low value for this parameter may, for instance, result in a mixture of components and cause low separability (Hassani et al., 2015). There is no consensus in the SSA literature regarding the determination of the parameter *L*. For example, Golyandina (2010) suggests utilizing values between 2 and N/2 as an appropriate approach for series decomposition. Hassani et al. (2011) recommend a value within the range of 2 and (N − 1)/2. Meanwhile, Hassani et al. (2009) describe that optimal separability occurs when $L = [N + 1]/2$.

Singular Value Decomposition:

[5] A Hankel-type matrix has the elements of the opposite diagonals *i* +*j* equal to a content value. For the SSA method, both the matrix columns and rows are subseries of the original series.

The subsequent stage of the process entails the decomposition of the matrix X into a sequence of rank-1 orthogonal matrices, which are referred to as elementary matrices. Furthermore, the matrix S can be expressed as the product of matrix X and its transpose, denoted as $S = \mathbf{X}\boldsymbol{X}^T$. Therefore, a singular value decomposition (SVD) is conducted.

In this manner, the eigenvalues of S are arranged in descending order, with λ_1 ≥ λ_2 ≥ ... ≥ $\lambda_L \geq \geq 0$. The system of orthogonal eigenvectors corresponding to the eigenvalues of S (U_1, U_2, ..., , U_L) is also a consequence of this process. Given that d represents the rank of X and its corresponding eigenvector is defined as $V_i = \boldsymbol{X}^T U_i/(\lambda_i)^{\frac{1}{2}}$, where i ∈ [1, d], it is established that d is equal to the minimum value of L and K. The matrix can be expanded through singular value decomposition (Golyandina et al., 2018), which can be expressed as Equation 4:

$$X = \sum_{i=1}^{d} (\lambda_i)^{\frac{1}{2}} U_i V_i^T = E_1, E_2, \ldots, E_L \quad (4)$$

- $\{E_l\}_{l=1}^{L} = (\lambda_i)^{\frac{1}{2}} U_i V_i^T$ are called elementary matrices.
- $(\lambda_i)^{\frac{1}{2}}$, i = 1, ..., L are the singular values, $\{(\lambda_i)^{\frac{1}{2}},\}_{l=1}^{L}$ is called the singular spectrum of X
- $\{U_i\}_{l=1}^{L}$ are the left singular vectors of the trajectory matrix X, also known as

empirical orthogonal functions (EOF).

- $\{U_i\}_{l=1}^{L}$ are the left singular vectors of the trajectory matrix X, also known as empirical orthogonal functions (EOF).
- $\{V_i\}_{l=1}^{L}$ are the right singular vectors of the trajectory matrix X, also known as factor vectors.

The collection $(\lambda_i)^{\frac{1}{2}}, U_{i,}, V_i)$ is known as the i-th eigentriple in the SVD of the trajectory matrix X. The contribution of each component is estimated by the ratio of the singular values, given by $(\lambda_i)^{\frac{1}{2}}, U_{i,}/\sum_{i=1}^{d}(\lambda_i)^{\frac{1}{2}}$.

One characteristic of the SVD process relates to the properties of the directions determined by the eigenvectors. Thus, the first eigenvector determines the direction such that the variation of the projections of the lagged vectors in this direction is maximized (Golyandina et al., 2018).

Clustering:

The next step, known as clustering, involves dividing the elementary matrices into several groups and summing the matrices within the same group. This process involves considering the sequence $\{E_l\}_{l=1}^{L}$ of elementary matrices from the decomposition. The

purpose is to separate additive components of the time series, thereby reducing the number of elementary matrices in the SVD of the trajectory matrix X (Hassani et al., 2009).

There are different techniques for grouping the matrices into $m \leq d$ groups. For example, clustering can be done using hierarchical clustering, by frequency, or even through graphical analysis of the obtained singular vectors. The result of this process is given by the equation 5 :

$$\boldsymbol{X} = \boldsymbol{X}_{I1} + \boldsymbol{X}_{I2} + \cdots + \boldsymbol{X}_{Im} \quad (5)$$

If the matrices $\boldsymbol{X}_{I1}$ and $\boldsymbol{X}_{I2}$are close to a Hankel form, it is possible to transform them into approximately separable time series $X^{(1)}\ and\ X^{(2)}$where the matrices are close to $\boldsymbol{X}_{I1}$ and $\boldsymbol{X}_{I2}$, respectively. This is a desirable property at this stage of the process (Hassani et al., 2011).

Diagonal Averaging:

The goal of diagonal averaging is to transform a matrix in the form of a Hankel matrix into a time series that is an additive component of the original series. If z_{ij} represents an element of Z, then the k-th term of the resulting series is obtained by averaging zz_{ij}over all i,j such that $i + j = k + 1$. This procedure is called diagonal averaging or Hankelization of the matrix Z.

The calculation of the k-th component of X from SSA is performed through diagonal averaging, following the rule:

$$\widetilde{x}_k = \begin{cases} \sum_{m=1}^{k} \frac{x_{m,k-m+1}}{k} & \text{if } 1 \leq k < L^*, \\ \sum_{m=1}^{L^*} \frac{x_{m,k-m+1}}{L^*} & \text{if } L^* \leq k < K^*, \\ \sum_{m=k-K^*+1}^{N-K^*+1} \frac{x_{m,m-m+1}}{N-k+1} & \text{if } K^* \leq k < N. \end{cases} \quad (6)$$

Let H be the process of transforming the matrices obtained from the clustering process into Hankel-type matrices, such that:

$$X = \widetilde{X_1} + \widetilde{X_2} + \cdots + \widetilde{X_m} \quad (7)$$

In this sense, $\widetilde{\boldsymbol{X}_{Is}} = H\ \widetilde{\boldsymbol{X}}$ is the result of the Hankelization process. Thus, an appropriate clustering makes the terms $\boldsymbol{X}_{Is}$approximately Hankel matrices, and the contribution can be evaluated through the ratio of the singular values.

Moreover, each trajectory matrix $\boldsymbol{X}_{Is}$obtained corresponds to a time series. Thus, the original series can be rewritten as a linear combination of the m estimated elemental series.

With the decomposed series, it will finally be used as input for a forecast model.

2.6 Regression Tree

The Regression Tree model is a machine learning approach widely used in time series forecasting. This model is particularly effective in identifying nonlinear relationships between explanatory variables and the target variable, segmenting the data into homogeneous subsets through hierarchical decision rules. Unlike traditional linear models, regression trees do not impose strict assumptions on data distribution, making them a robust alternative for time series with complex patterns (Rady et al, 2021).

The Regression Tree model is structured as follows:

- Decision nodes: Points where data is split based on a cutoff criterion;
- Branches: Connections between nodes, representing different possible trajectories in data segmentation;
- Terminal nodes: They represent the values predicted by the model, generally calculated as the average of the values of the response variable within each partition.

The algorithm seeks to identify the optimal split points that segment the data into more homogeneous groups regarding the target variable. This partitioning process is iterative and continues until stopping criteria are met. Each split is determined by maximizing variance reduction, as show in Equation 8, ensuring that the final partitions capture meaningful patterns within the data (Breiman *et al, 2017*).

$$\Delta VAR = VAR(T) - \left(\frac{|T_L|}{|T|}VAR(T_L) + \frac{|T_R|}{|T|}VAR(T_R)\right) (8)$$

Where:

- $VAR(T_L)$ is the varience in the original node before de slitting;
- $VAR(T_L)$ and $VAR(T_R)$ represent the variances in the resulting left and right nodes, respectively;
- $|T_L|$ and $|T_R|$ are the sizes of the generated subsets.

2.7 Multi-Layer Perceptron

Multi-Layer Perceptron (MLP) isa deep learning model widely used in time series forecasting. The MLP is an artificial neural network composed of multiple interconnected layers of neurons, capable of capturing complex relationships between variables by adjusting synaptic weights during training. This model is particularly useful for time series forecasting as it can learn nonlinear patterns and represent underlying dependencies in the data (Liu *et al, 2022*).

The MLP structure consists of three main types of layers: the input layer, which receives the predictor variables; one or more hidden layers, responsible for processing information and learning abstract representations; and the output layer, which generates the model's predictions. Each neuron in the hidden layers applies a linear transformation followed

by a nonlinear activation function, allowing the network to capture complex interactions in the data (Ciaburro *et al, 2017*).

The model's operation is mathematically described in Equation 9. Given an input *x,* the neurons in the hidden layer perform the transformation:

$$h_j = \sigma(W_J x + b_j) \quad (9)$$

Where:

- W_J represents the synaptic weights associated with neuron *j;*
- x represents the input;
- b_j is the bias term;
- σ is the activation function.

The model's output is then obtained through a linear combination of the last hidden layer's neurons, as described in Equation 10:

$$y = W_0 h + b_0 \quad (10)$$

Where:

- *y* represents the output;
- W_0 represents the weights matrix associates with output layer;
- h represents the Vector of activations of neurons in the last hidden layer;
- b_0 represents output layer bias vector.

The model was trained using the backpropagation algorithm with is responsible fort adjusting the network's synaptic weights by minimizing the error between model predictions and actual values. To evaluate the model's performance, the dataset was split into training and test sets, ensuring that predictions were validated on unseen data during model fitting (Ciaburro *et al, 2017*)..

The application of the MLP model in time series forecasting enabled the efficient modeling of complex patterns, capturing variable interactions and temporal dependencies. The neural network's ability to learn nonlinear relationships without requiring manual specification of interactions makes this approach a powerful tool for predictive analysis (Liu *et al, 2022*).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (11)$$

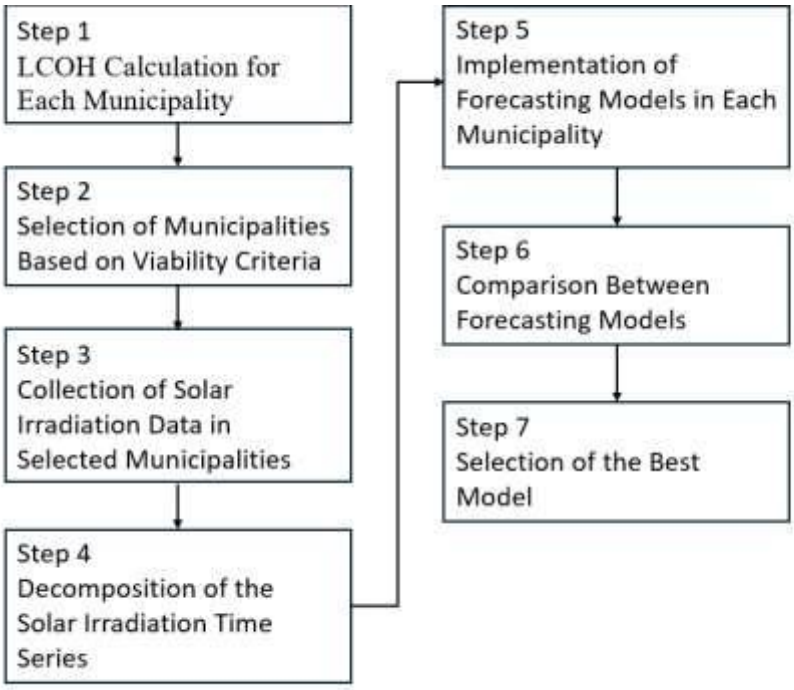
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (12)$$

To evaluate the accuracy of the models analyzed, two widely used statistical metrics were employed: Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), the mathematical equations are presented by equations 11 and 12, respectively. These metrics are essential for quantifying the discrepancy between predicted and observed values, providing different perspectives on model accuracy.

3 **METHODOLOGY**

To meet the proposed objective, the present dissertation was organized into 7 steps, as represented in the flowchart in Figure 1.

Figure 1 - Flowchart of Methodology



Source: Author's own elaboration

3.1 LCOH Calculation for Each Municipality

The first step of this study consists of calculating the Levelized Cost of Hydrogen (LCOH) for each Brazilian municipality. This cost represents the minimum price at which hydrogen can be sold to cover all expenses involved in its production, including initial investments, operation, and plant maintenance over time.

For this analysis, a green hydrogen production plant is considered, comprising the following components: a photovoltaic power plant for electricity generation, a battery bank for storage and energy supply stability, an electrolyzer responsible for splitting water into hydrogen and oxygen, a compressor to pressurize the produced hydrogen, and a storage tank.

Hydrogen generation occurs through water electrolysis, requiring two main inputs: electricity from the photovoltaic plant and water. For each system component, both capital expenditures (CAPEX) and operational expenditures (OPEX) are considered, as presented in Table 1 (De Andrade et al., 2024).

Table 1 - Inputs parameters used for calculating LOCH

Parameter	Value
Solar PV price ((US$/Wp)	0.77
PV OPEX (US$)	0.5
Electrolyze CAPEX(US$/kW)	1000
Electrolyze OPEX(US$)	2
Water Cost (US$/$m^3$)	2.43
Eletricity Cost (US$/kWh)	0.05
Compressor CAPEX(US$)	1,300
Compressor OPEX (US$)	0.8
Storage CAPEX (US$/N$m^3$)	63
Storage OPEX (US$/day)	0.5
Battery Bank CAPEX (US$/kWh)	415.2
Discount Rate (%)	10.5

Source: Author's own elaboration

LCOH will be calculated considering the energy generated by the photovoltaic plant throughout its lifetime, operational costs associated with equipment maintenance, and the efficiency of electrolysis in the hydrogen production process. The calculation method accounts for cash flow discounting over time, applying the discount rate to determine the final levelized cost.

3.2 Selection of Municipalities Based on Viability Criteria

The second step consists of selecting municipalities based on financial viability criteria for green hydrogen production. For comparative analysis, two municipalities are selected: one with the highest economic viability and one with the lowest viability. This process helps understand structural differences between scenarios and assess the impact of variables on project feasibility determination.

3.3 Collection of Solar Irradiation Data in Selected Municipalities

For each selected municipality, an hourly time series of solar irradiation (kW/m^2) will be constructed. The data were obtained from the Prediction of Worldwide Energy Resources (POWER) program, developed by NASA, covering the period from 2022 to 2023.The data were extracted and processed using the R programming language with the "nasapower" package (Sparks et al., 2018), which enables direct access to historical solar irradiation data. After that, the data is organizing into time series and prepared for subsequent analyses.

3.4 Decomposition of the Solar Irradiation Time Series

With the obtained time series, decomposition will be performed using the Singular Spectrum Analysis (SSA) technique. This method allows the separation of the series components into:

- Trend: represents long-term variation in solar irradiation.
- Cyclic Patterns: periodic fluctuations associated with seasonality and climatic variations.
- Residual Component: random errors without clear patterns.

The graphical analysis is use to the identification of the components.

3.5 Implementation of Forecasting Models in Each Municipality

Based on the decomposition results, the identified trend, cyclic, and residual components for each municipality will be used as explanatory variables in forecasting models, as well as, the lags of the the solar irradiance for t (1, 2,…,24) and the dummies identifying the hours with each observation is correspond to. The dependent variable in these models is

solar irradiation. To improve model accuracy, the rolling window approach will be use as a training method (Aparma, 2018; Brown *et al, 2024*).

3.6 Comparison Between Forecasting Models

After estimating the models, a ARIMA (autoregressive integrated moving average) model will be considered as a benchmarking reference and compared with the other models. The models will be evaluated based on error metrics such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Additionally, the predicted values will be analyzed against observed values to verify how closely they match.

3.7 Selection of the best model

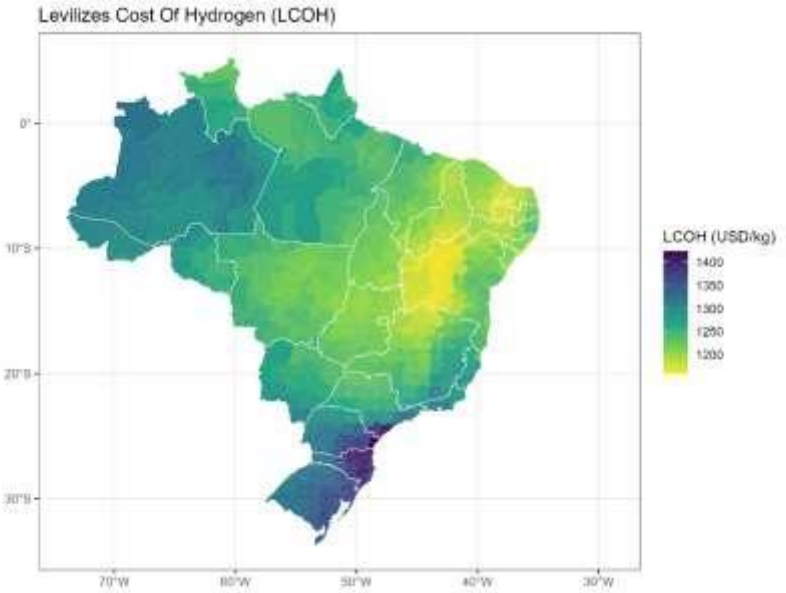
Following the accuracy analysis, the model with the most favorable metrics will be selected. The chosen model will be used for future predictions of solar energy generation

4 RESULTS

4.1 LCOH

This section presents the results of levelized cost of hydrogen (LCOH) in Brazilian municipalities. The analysis revealed significant disparities, as shown in Figure 2, highlighting the influence of regional factors on the economic feasibility of green hydrogen production. These findings are crucial for guiding public policies and strategic investments, as hydrogen costs are one of the main determinants of its competitiveness compared to other energy sources.

Figure 2 - Levelized cost of hydrogen for Brazilian municipalities

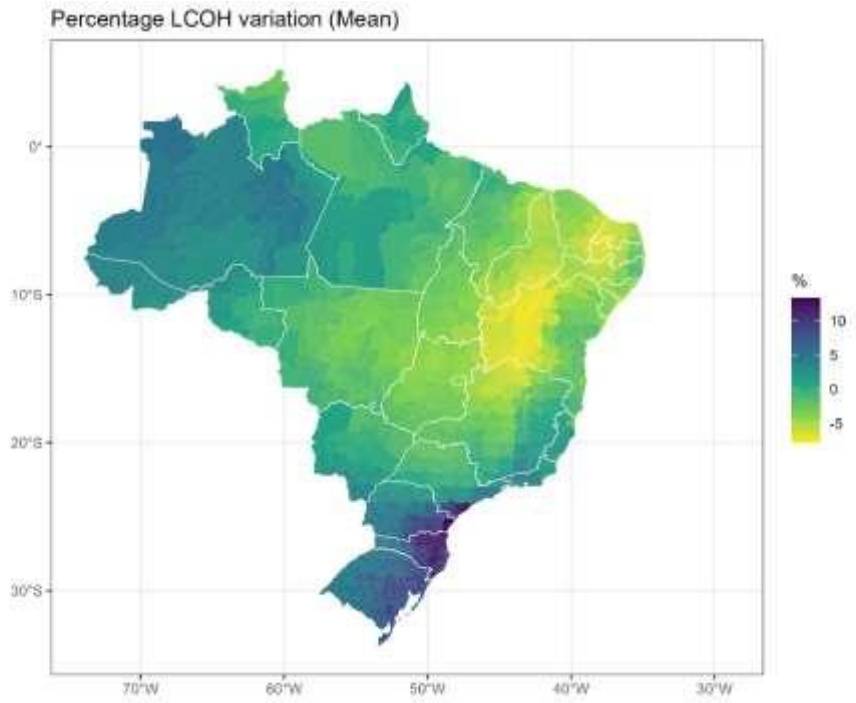


Source: Author's own elaboration

The lowest LCOH value was recorded in Xique-Xique, Bahia (BA), a municipality located in the central region of the state of Bahia (BA), approximately 550 km from Salvador, reaching R$ 1,159.22 (USD/kg) This result suggests that the municipality has highly favorable conditions for green hydrogen production, which can be attributed to the abundant availability of renewable energy resources. Xique-Xique's geographical location, in a region with high solar radiation levels and may enable high efficiency in converting renewable energy into electricity, thereby reducing electrolysis costs. Additionally, the lower land costs in the region may contribute to reducing initial investments (CAPEX), while water availability may positively impact the efficiency of the electrolysis process, depending on the quality and treatment requirements of the resource.

On the other hand, Antonina, Paraná (PR) a municipality located in the littoral of Paraná (PR), approximately 120 km south of Curitiba, near the southern coast of Brazil, presented the highest LCOH in the study, reaching R$ 1,424.32 (USD/kg). This value represents a significant increase compared to the national average, 1,256.99 (USD/kg). Possible factor contributing to this higher cost is the lower availability renewable energy, Although Paraná has a significant renewable energy matrix based on hydroelectric power, it may face challenges int the solar generation, affecting the electricity cost used in electrolysis.

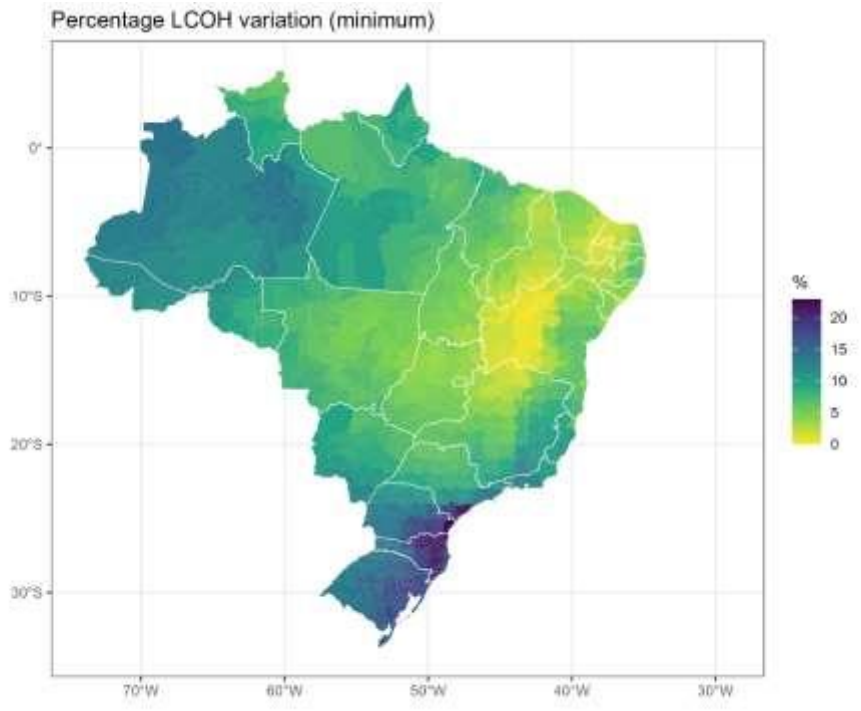
Figure 3 - Percentage LCOH variation (Mean)



Source: Author's own elaboration

The percentage variation of LCOH compared to the national average, Figure 3, further reinforces these regional differences. Xique-Xique (BA) registers a -7.78% reduction, highlighting its competitive potential for green hydrogen production. This favorable performance may position it as a strategic hub for sector investments, whether to supply the domestic market or for export, especially given the growing global demand for green hydrogen. In contrast, Antonina (PR) shows an increase of 13.31% above the national average, suggesting that, without incentives and technological advancements, the economic feasibility of hydrogen production in this region may be challenging.

Figure 4 - Percentage LCOH variation (minimum)



Source: Author's own elaboration

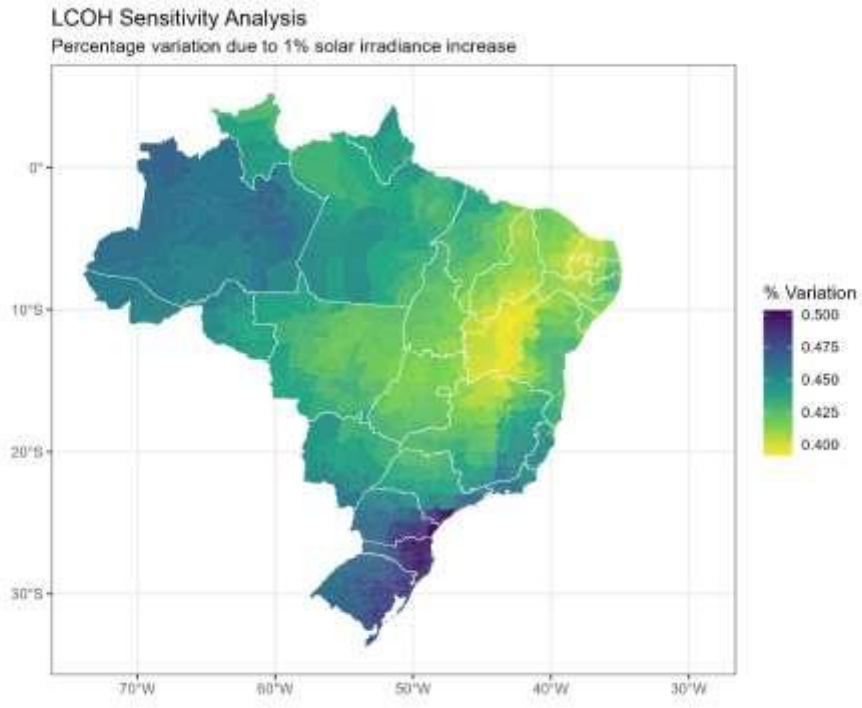
The percentage variation of LCOH compared to lowest municipality, as portrayed in Figure 4, reinforcing the importance of strategic site selection for green hydrogen projects. When comparing Antonina (PR) to the lowest identified LCOH value, an increase of 22.87% is observed, reinforcing the importance of strategic site selection for green hydrogen projects. This contrast between municipalities, as portrayed in, indicates that the economic success of production is strongly linked to energy conversion efficiency and costs associated with the production chain, including electrolysis and storage.

In addition to geographical and infrastructural factors, public policies and financial incentives play a crucial role in reducing LCOH. Countries that have implemented subsidy mechanisms or tax exemptions for green hydrogen production have significantly lowered costs, making hydrogen more competitive against fossil fuels. In Brazil, initiatives such as auctions specifically for green hydrogen, differentiated financing lines, and investments in research and innovation may be fundamental to improving economic viability, particularly in regions where costs are higher.

These findings emphasize the influence of multiple factors on the feasibility of green hydrogen in Brazil. Northeastern regions, exemplified by Xique-Xique, offer competitive advantages due to their abundance of renewable resources and lower operating costs, making them promising candidates for sector expansion. Meanwhile, some areas in the South face additional challenges, such as lower renewable energy efficiency, which may require specific

policies to make production viable in these areas. Therefore, for Brazil to become a global leader in green hydrogen production, it will be essential to develop regionally tailored strategies that maximize each location's competitive advantages and minimize the specific challenges faced by less favored regions.

Figure 5 - LCOH Sensitivity Analysis



Source: Author's own elaboration

Finally, it was conducted a sensitivity analysis, Figure 5, considered a 1% increase in average solar irradiation, and evaluated the impact of this variation on the cost structure of Green Hydrogen (LCOH) in Brazilian municipalities. The results indicate that, on average, the percentage variation in LCOH was -0.437%, meaning there was a cost reduction with the increase in solar energy availability. This relationship was expected, as higher solar irradiation improves electrolysis efficiency by reducing electricity costs, one of the main components of LCOH. The data also show a dispersion in impacts among the analyzed municipalities. While some recorded more significant reductions (with a minimum of -0.392% and a maximum of -0.503%), others experienced smaller impacts. The regional results reveal that the South had the highest average percentage variation in LCOH (-0.451%), indicating that municipalities in this region are more sensitive to changes in solar irradiation. The Southeast/Central-West had the lowest average variation (-0.419.) The Northeast/ North showed intermediate values (-0.437%), which may be associated with the already high availability of solar resources.

This analysis highlights the importance of accurate forecasting models for solar irradiation. in the sense that, small inaccuracies can generate significant differences in LCOH estimates, impacting decisions regarding the location of production plants and investment

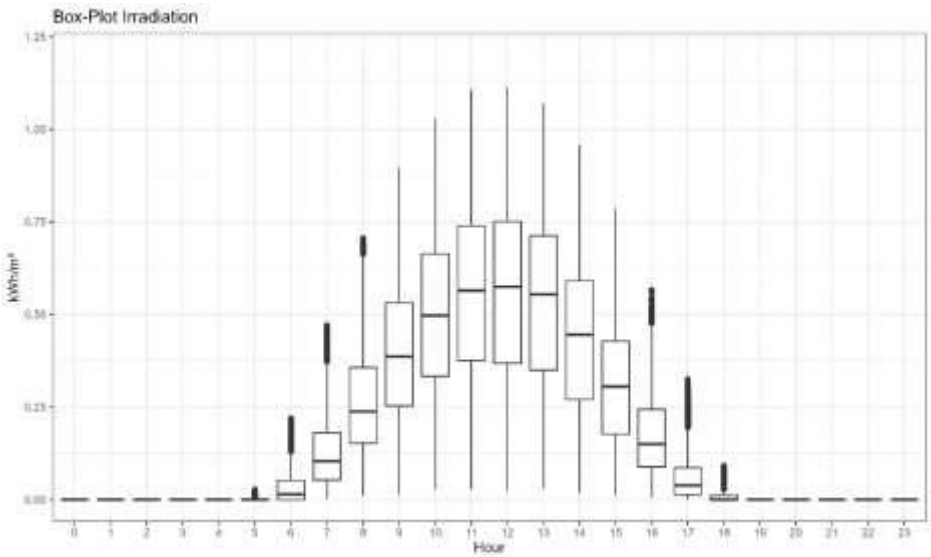
attraction. The observed regional variability reinforces that simplistic model relying solely on historical averages of solar irradiation may not be sufficient for precise LCOH calculation. Furthermore, errors in irradiation forecasts can lead to suboptimal investment decisions, affecting financial planning and operational efficiency. Overestimating solar potential may result in financial losses due to lower-than-expected energy yields, while underestimating it could deter investments in viable locations.

4.2 Results for the municipality of Xique- Xique (BA)

4.2.1 Descriptive Analysis

In the descriptive analysis of hourly solar irradiation reports significant patterns related to the daily cycle, as shown in Figure 6, and seasonal variations of throughout the four seasons of the year, such as Winter, Spring, Summer, and Fall,. It was observed that during the early hours of the morning (12 a.m. to 5 a.m.), irradiation is consistently null across all seasons, reflecting the absence of solar radiation. Starting at 6 a.m., there is a progressive increase in irradiation, with peaks generally occurring at 11 a.m. in all seasons. Notice that, by peak sun hours it is defined as one hour in which the intensity of solar irradiance is higher than the rest of the day, normally achieved by 11 a.m. After the peak, irradiation values gradually decrease, stabilizing at zero again after 6 p.m.

Figure 6 - irradiation throughout the year (Xique-Xique (BA))

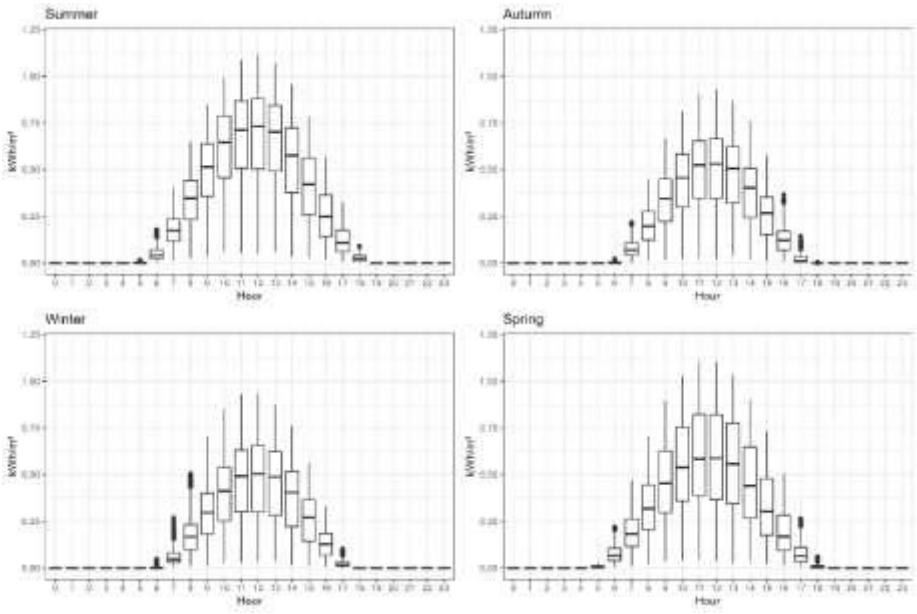


Source: Author's own elaboration

In each season there is a difference value at the peak hours, as show in Figure 7. For example, during the Winter, the average irradiation at the peak was 0.88 kW/m², with a maximum of 1.07 kW/m² and a standard deviation of 0.12 kW/m², indicating lower variability compared to the other seasons. In comparison, at the Fall season, the average irradiation at the peak was slightly lower, reaching 0.80 kW/m², with a maximum of 1.05 kW/m² and a standard deviation of 0.13 kW/m². Heading towards Spring, the highest average values of

irradiation were observed, reaching 0.92 kW/m², with a maximum of 1.11 kW/m² and higher variability, represented by a standard deviation of 0.18 kW/m². Summer, in turn, exhibited an average peak irradiation of 0.88 kW/m², a maximum of 1.10 kW/m², and a standard deviation of 0.17 kW/m².

Figure 7 - Solar irradiation throughout the seasons (Xique-Xique (BA)



Source: Author's own elaboration

The comparative analysis among the seasons shows that Summer and Spring recorded the highest average and maximum irradiation values, accompanied by greater hourly variability. During Winter, the lowest irradiation values at peak hours were recorded, reflecting shorter days and lower solar intensity typical of this season. Fall, on the other hand, exhibited intermediate values, suggesting a gradual transition to a period of lower solar incidence.

In summary, the analyzed data reveal a clear pattern of hourly solar irradiation variation influenced by both the daily cycle and seasonal changes. Summer and Spring stand out for their higher intensity and variability of solar radiation, while Winter is characterized by more modest and uniform irradiation values throughout the day.

4.2.2 Singular Spectrum Analysis

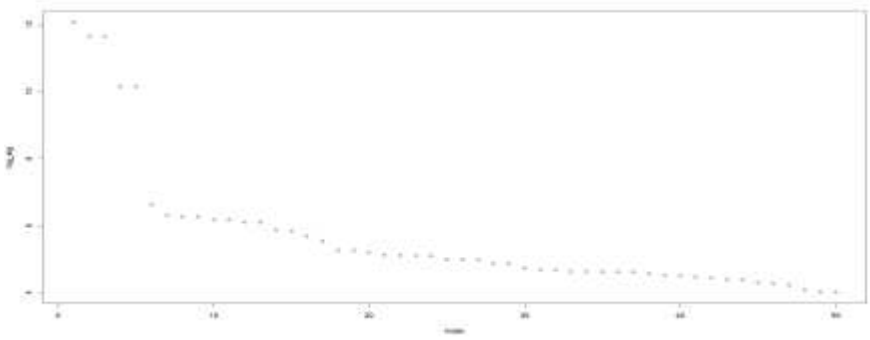
This section presents the results of the Singular Spectrum Analysis (SSA) model implementation. Regarding the choice of parameter L (Golyandina et al., 2013; Hassani et al., 2018); Yang et al, 2021), considering that there are indications of seasonality in the analyzed

series, a value of 4,380 was chosen, corresponding to 365×12 This selection aims to capture both daily and monthly seasonality.

One of the approaches for defining the SSA groupings is based on the analysis of graphical components, such as eigenvalue plots, eigenvectors, eigenvalue pairs, and the weighted correlation matrix (W) (Golyandina et al., 2013; Hassani et al., 2018). This methodology allows distinguishing which components can be classified as trend and cyclical and which represent noise.

Based on Logarithm of eigenvalue, Figure 8, it is possible to determine which eigentriples are associated with the trend component. To do so, the largest eigenvalues should be examined, as well as the distance between them. In this context, it can be inferred that the first eigenvalue is related to the trend, as it not only has the highest value but also shows a significant separation from the others.

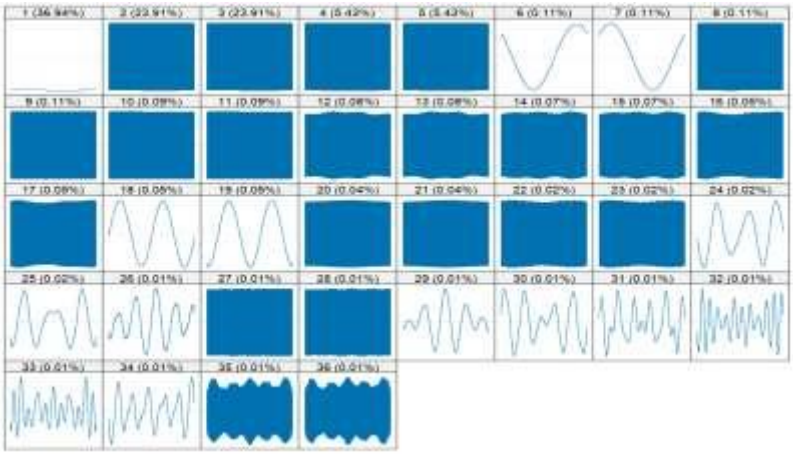
Figure 8 - Logarithm of eigenvalue



Source: Author's own elaboration

Regarding the eigenvalues, as illustrated in Figure 9 it is essential to emphasize that their relative contribution indicates not only which ones exhibit slower variation but also their impact on the total variance of the series. The first eigenvector has the highest contribution, accounting for 36.94%, followed by the second eigenvalue at 23.91%. A progressive decline in the significance of the remaining eigenvalues is observed, with the relative contribution dropping below 1% from the sixth eigenvalue onward. Additionally, from the tenth eigenvalue, the contribution falls below 0.1%. These insights are crucial for reconstructing the components as they determine the relevance of each component in the original series' composition.

Figure 9 - Eigenvectors and their relative contributions



Source: Author's own elaboration

The analysis of eigenvector pairs, Figure 10, helps identify the components associated with the harmonic part of the series based on the geometric patterns displayed. The first pair exhibits low variation due to the influence of the first component, which is linked to the trend. it is observed that pair 3x4 exhibit more regular shape, with patterns resembling smooth sinusoidal cycles. This suggests the presence of a dominant cyclical structure associated with the daily variation in solar irradiation. This behavior aligns with the expectation that irradiation is null at night, gradually increases throughout the day, reaches a peak around noon, and decreases until sunset. The approximately elliptical or circular shape observed, for example in the pairs 4×5, and 5×7 indicates a relationship between these components and a predominant faster cycle.

As we move further into the matrix, some pairs display more complex patterns, with multi-sided polygonal shapes, pairs 8x9 and 27x28, for example. These patterns indicate the presence of variations that are not limited to the daily cycle but also include medium-term oscillations, possibly associated with intra-seasonal and monthly variations. This is consistent with the observation that solar irradiation variability tends to be higher during summer months, when there is greater radiation availability and increased atmospheric turbulence, and lower during winter months, when days are shorter and irradiation remains more stable throughout the day.

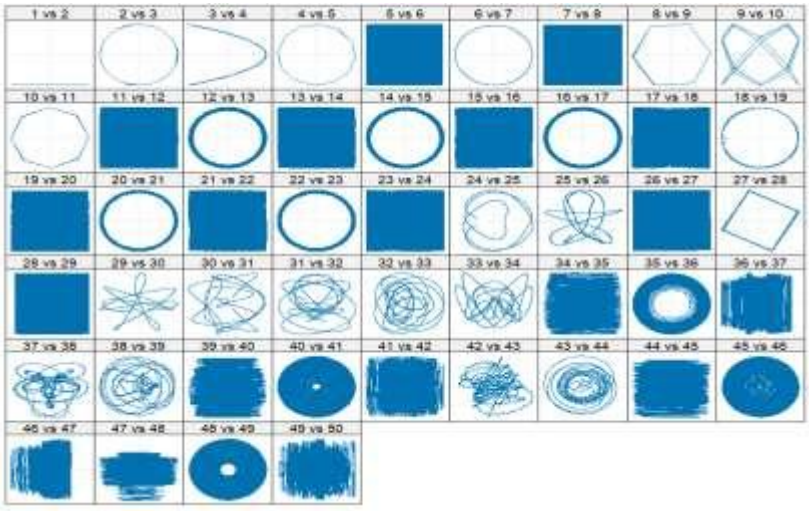
Additionally, some of the more advanced pairs of eigenvectors exhibit highly irregular patterns or nearly homogeneous blue-filled areas, for instance pairs 28x29 to 49x50. These patterns indicate the presence of noise components or high-frequency fluctuations, which may be related to daily atmospheric variability, such as cloud cover, short-term meteorological oscillations, and other sources of interference that do not follow a well-defined pattern.

Another relevant aspect is that the polygons formed in certain pairs of eigenvectors reflect the periodicity associated with different time scales. Polygons with fewer sides may be linked to lower-frequency seasonal oscillations, such as annual variations in irradiation due to seasonal changes. On the other hand, more complex and enclosed shapes may be associated with short-term phenomena, such as rapid meteorological changes.

Therefore, the decomposition of the time series through SSA enables the identification of multiple seasonal components in solar irradiation, ranging from daily variability to seasonal variations throughout the year. The analysis of the pairs of eigenvectors and their respective

graphical representations suggests the presence of well-defined cyclical patterns, reinforcing the influence of astronomical and atmospheric factors on the dynamics of solar irradiation.

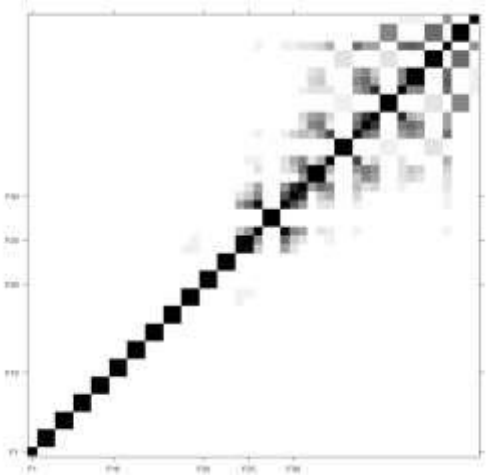
Figure 10 - Pairs of Eigenvectors



Source: Author's own elaboration

Figure 11 presents the weighted correlation matrix (W), which aids in classifying the components to be used in the series reconstruction. This matrix displays correlations among the factors on a continuous scale, where darker shades indicate high correlations (close to 1), while lighter shades represent low correlations (close to 0). It can be observed that the first factor (F1) has a low correlation with the others, suggesting high separability from the remaining components. Furthermore, factors 2 through 20 show strong correlations not just among themselves, there is always a correlation between 2 and 3, 4 and 5, 6 and 7 and so on, more or less halfway. This corroborates the analysis of paired eigenvectors given that those with the same seasonalities form a polygon and those with different seasonalities do not form a polygon (e.g.: 3 and 4, 5 and 6, 7 and 8). Beyond this range, a progressive decline in correlation is evident, with the subsequent factors becoming more correlated with each other, potentially indicating the presence of noise.

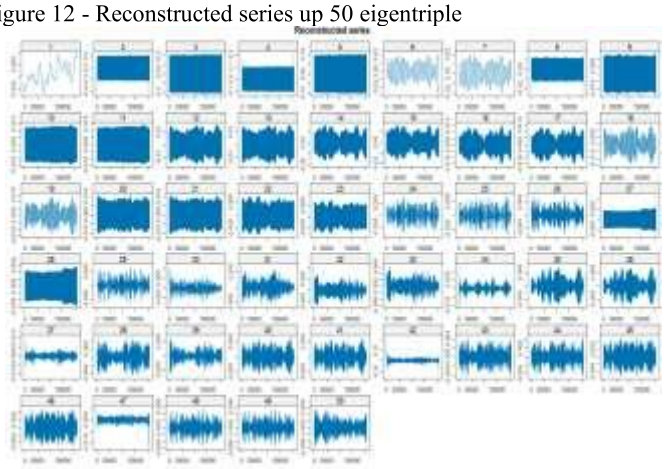
Figure 11 - W Correlation Matrix



Source: Author's own elaboration

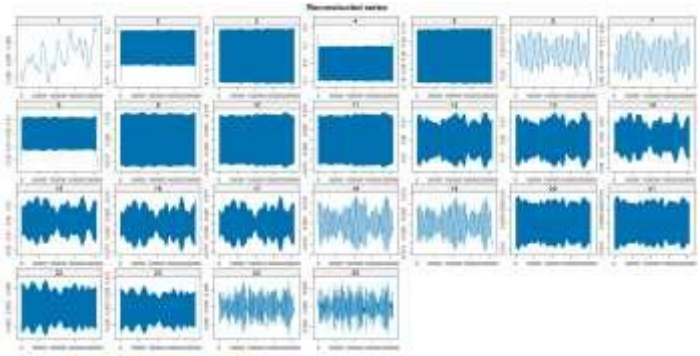
Figures 12 and 13 present the reconstructed series graphs obtained in the final stage of SSA application, offering further insight into which eigentriples should be used to reconstruct each component of the series. For instance, series 6 and 7 exhibit cyclical behavior around a mean, suggesting they belong to the signal component. In contrast, series 24 and 25, characterized by high frequency and large variance, can be classified as part of the noise component.

Figure 12 - Reconstructed series up 50 eigentriple



Source: Author's own elaboration

Figure 13 - Reconstructed series up 25 eigentriple



Source: Author's own elaboration

It is crucial to highlight that the selection of eigentriples for each component must integrate multiple analytical metrics. Identifying the eigentriple corresponding to the trend, for example, involves factors such as having the highest log-eigenvalue, the highest relative contribution, and a first factor with low correlation with the others. Similarly, the selection of eigentriples associated with noise is determined by the lowest log-eigenvalue, low relative contribution, absence of identifiable patterns in eigenvector pairs, high correlation among factors, and characteristics such as high frequency and large variance in the reconstructed series.

At the conclusion of the graphical analysis of the SSA estimation results, the following classification is established: the first eigentriple represents the trend component, eigentriples 2 through 22 correspond to the cyclic component, and the remaining eigentriples are classified as noise.

4.2.3 Forecast Model

This section presents and analyzes the results obtained from the forecasting model. Regarding the forecasting horizon, a value of h = 24 was chosen, representing a complete daily cycle of solar irradiation. The primary objective of selecting this horizon is to capture fluctuations throughout the day, allowing for a detailed analysis of intra-day variations. This aspect is crucial, as solar irradiation exhibits dynamic and non-stationary behavior influenced by meteorological and seasonal factors. By using a 24-hour period, the aim is to gain a better understanding of short-term patterns, which can be critical for applications that rely on accurate forecasts, such as solar energy management.

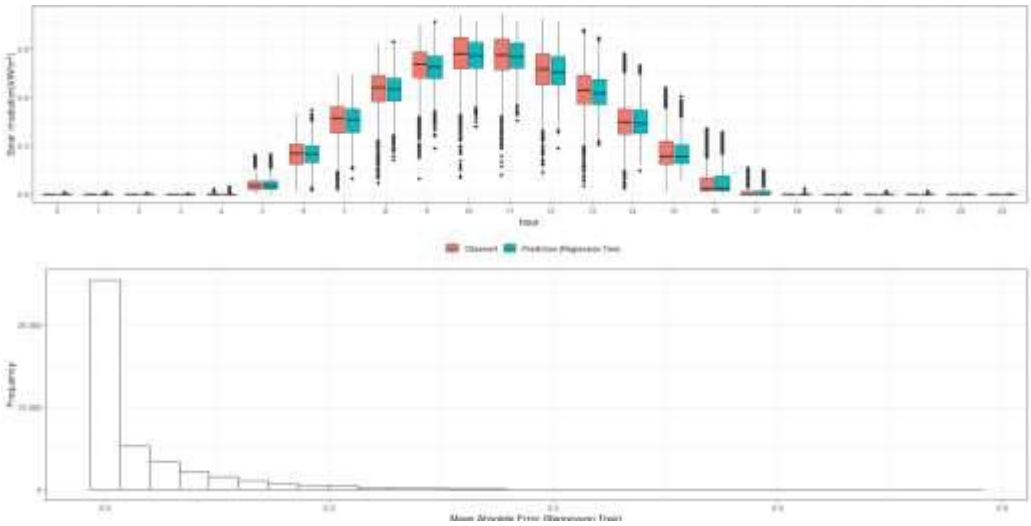
The results obtained for each model are summarized in Table 1. Firstly, it was observed that the benchmarking model performed worse compared to the other models. Analyzing the metric values, it is observed that, in both evaluations, the SSA + MPL model outperforms the SSA + Regression Tree model. This result suggests that combining the SSA decomposition technique with ANN enables the model to capture more relevant patterns in the time series, effectively reducing forecasting errors.

Table 2 - Error Metrics (Xique-Xique(BA))

Model	*RMSE*	*MAPE*
ARIMA	3.13	0.756
SSA + MLP	0,0004	0,012
SSA + Regression Tree	0,003	0,029

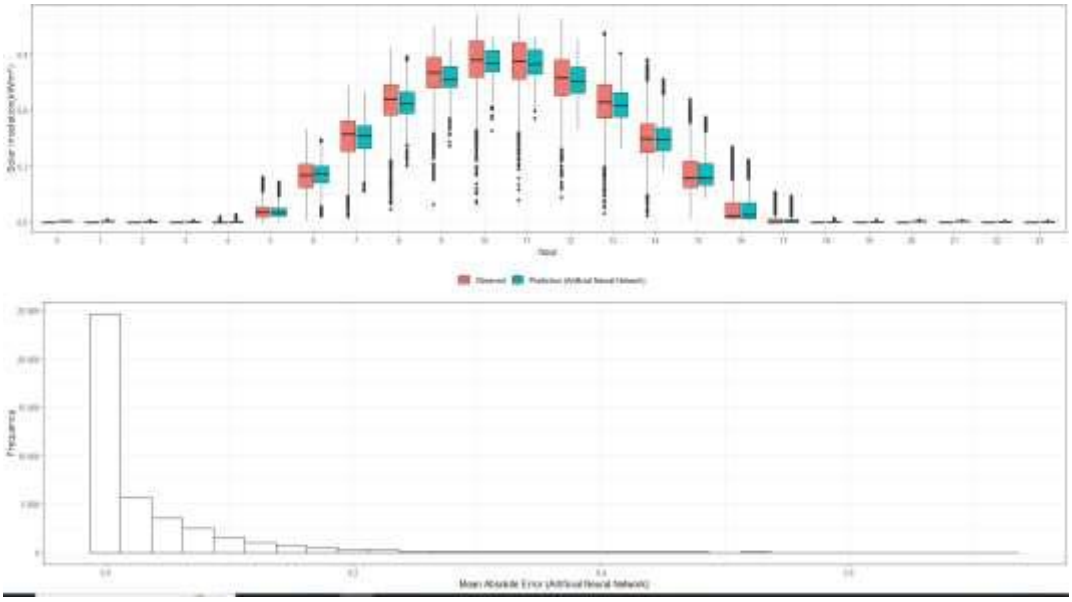
Source: Author's own elaboration

Figure 14 - Forecast Model Accuracy (SSA+MPL)



Source: Author's own elaboration

Figure 15 - Forecast Model Accuracy (SSA+ Regression Three)



Source: Author's own elaboration

Figures 14 and 15 illustrate the comparison between observed and predicted values obtained from Artificial Neural Network (ANN) and the Regression Tree models, respectively. In both cases, the first chart presents the distribution of observed (red) and predicted (blue) values throughout the day using boxplots. As expected, solar irradiation remains close to zero during nighttime hours (12 AM to 5 AM and after 6 PM), increases from 6 AM, reaches its peak between 10 AM and 2 PM, and gradually decreases after 3 PM. Analyzing the Regression Tree (Figure 14), it is evident that the predictions closely follow the observed values, particularly during peak irradiation hours. The median of the predictions aligns well with the observed values, indicating that the model captures the general trend effectively. However, there is higher dispersion during transition periods (early morning and late afternoon), where abrupt variations in irradiation occur. Additionally, some outliers are observed, especially during sunrise and sunset, which may be attributed to atmospheric conditions not captured by the model, such as cloud cover and sudden weather changes.

The second charts in Figures 14 and 15 present the frequency distribution of the Mean Absolute Error (MAE) for each model. In both cases, the errors are predominantly concentrated near zero, suggesting high predictive accuracy. However, the Regression Tree model exhibits a slightly longer right tail, indicating a higher occurrence of large errors compared to the ANN. These larger errors may stem from difficulties in capturing sudden variations in solar irradiation.

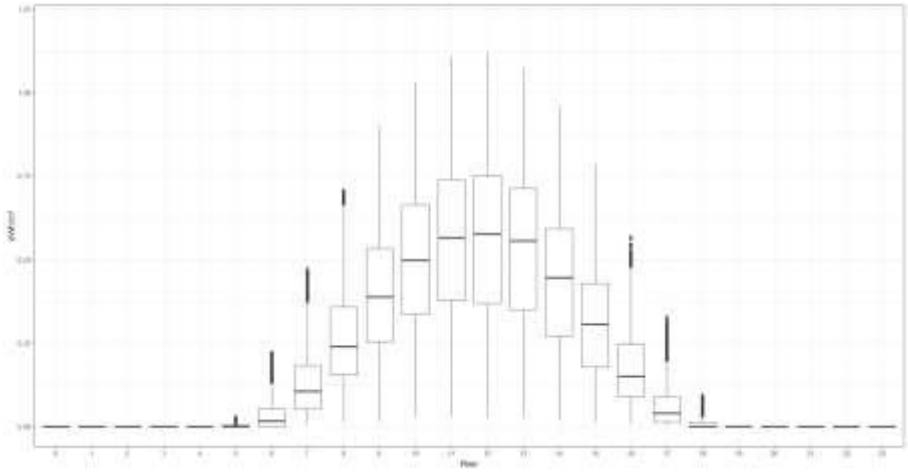
Overall, the ANN model demonstrates superior predictive performance, achieving lower RMSE and MAPE values while maintaining strong alignment between predicted and observed values. Additionally, another crucial factor to consider is computational efficiency.

4.3 Results for the municipality of Antonia (PR):

4.3.1 Descriptive Analysis

The descriptive analysis of hourly solar irradiation reveals significant patterns related to the daily cycle, as shown in Figure 16, and seasonal variations throughout the four seasons of the year: Winter, Spring, Summer, and Fall. It was observed that during the early hours of the morning (12 a.m. to 5 a.m.), irradiation is consistently null across all seasons, reflecting the absence of solar radiation. Starting at 6 a.m., there is a progressive increase in irradiation, with peaks generally occurring at 11 a.m. in all seasons.

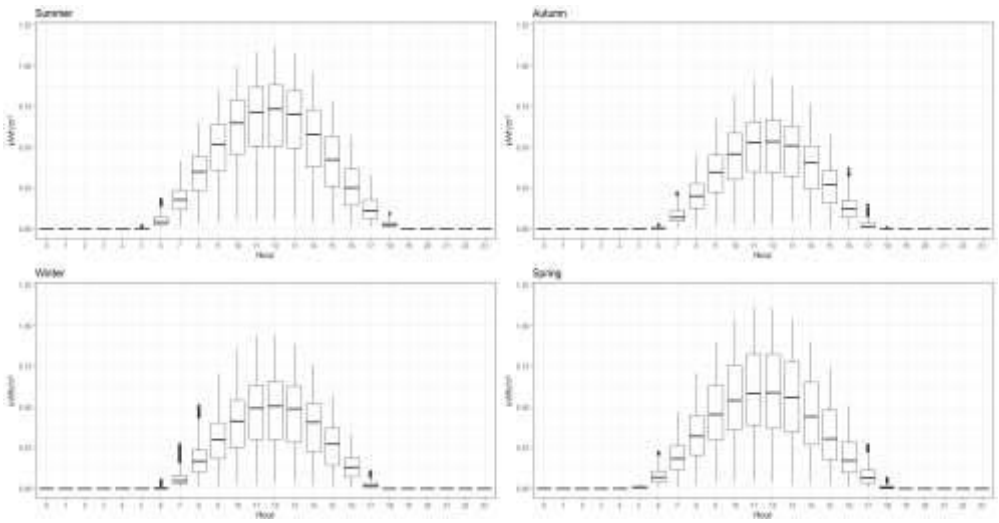
Figure 16 - irradiation throughout the year (Antonia (PR))



Source: Author's own elaboration

The concept of peak sun hours refers to the time when the intensity of solar irradiance is at its highest, typically around 11 a.m. After this peak, irradiation values gradually decrease, stabilizing at zero again after 6 p.m. Figure 17 illustrates the variations in irradiation values at peak hours across different seasons.

Figure 17 - Solar irradiation throughout the seasons (Antonia (PR))



Source: Author's own elaboration

The comparative analysis among the seasons shows that Summer and Spring recorded the highest average and maximum irradiation values, accompanied by greater hourly variability. During Winter, the lowest irradiation values at peak hours were recorded, reflecting shorter days and lower solar intensity typical of this season. Fall, on the other hand, exhibited intermediate values, suggesting a gradual transition to a period of lower solar incidence.

In summary, the analyzed data reveal a clear pattern of hourly solar irradiation variation influenced by both the daily cycle and seasonal changes. Summer and Spring stand

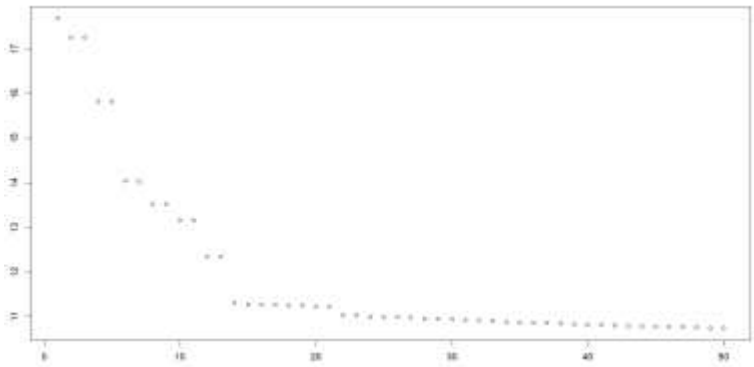
out for their higher intensity and variability of solar radiation, while Winter is characterized by more modest and uniform irradiation values throughout the day.

4.3.2 Singular Spectrum Analysis

This section presents the results of the Singular Spectrum Analysis (SSA) model implementation. Regarding the choice of parameter L(Golyandina et al., 2013; Hassani et al., 2018; Yang *et al*, 2021), considering that there are indications of seasonality in the analyzed series, a value of 4,380 was chosen, corresponding to 365×12 This selection aims to capture both daily and monthly seasonality.

Based on Logarithm of eigenvalue, Figure 18, it is possible to determine which eigentriples are associated with the trend component. To do so, the largest eigenvalues should be examined, as well as the distance between them. In this context, it can be inferred that the first eigenvalue is related to the trend, as it not only has the highest value but also shows a significant separation from the others.

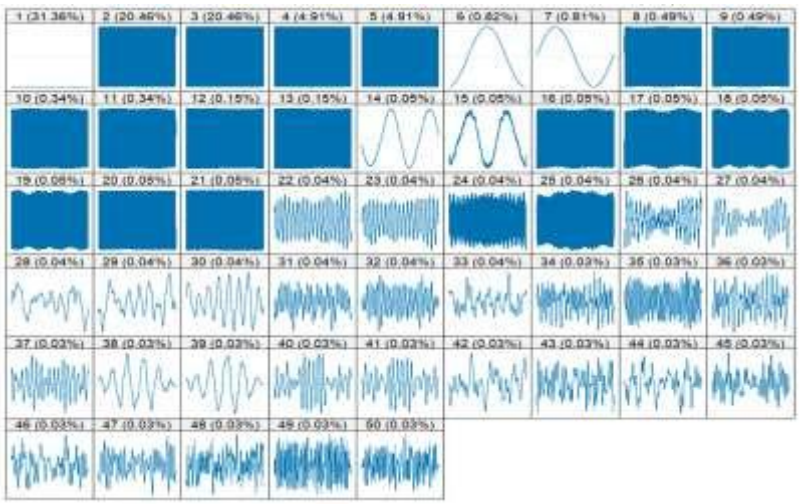
Figure 18 - Logarithm of eigenvalue



Source: Author's own elaboration

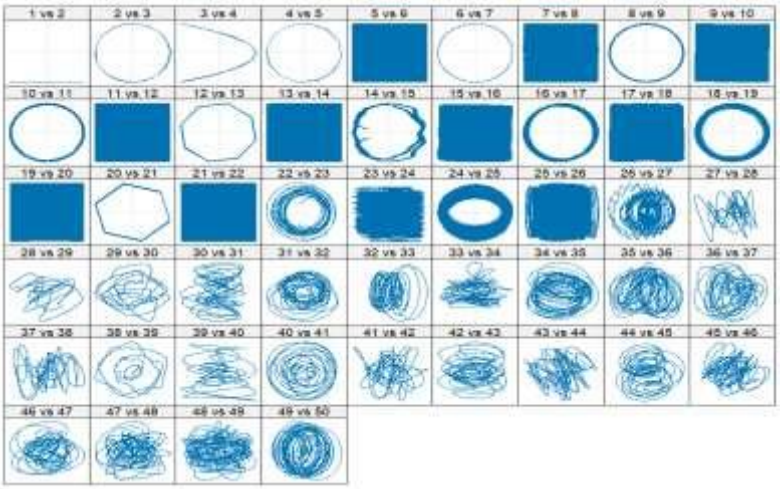
Regarding the eigenvalues, as illustrated in Figure 19, it is essential to emphasize that their relative contribution indicates not only which ones exhibit slower variation but also their impact on the total variance of the series. The first eigenvector has the highest contribution, accounting for 31.36%, followed by the second eigenvalue at 20.46%. A progressive decline in the significance of the remaining eigenvalues is observed, with the relative contribution dropping below 1% from the sixth eigenvalue onward. Additionally, from the fourteenth eigenvalue, the contribution falls below 0.1%. These insights are crucial for reconstructing the signal of the series, as they determine the relevance of each component in the original series' composition.

Figure 19 - Eigenvectors and their relative contributions



Source: Author's own elaboration

Figure 20 - Pairs of Eigenvectors



Source: Author's own elaboration

The analysis of eigenvector pairs, Figure 20, helps identify the components associated with the harmonic part of the series based on the geometric patterns displayed. The first pair exhibits low variation due to the influence of the first component, which is linked to the trend. it is observed that pair 3x4 exhibit more regular shape, with patterns resembling smooth sinusoidal cycles. This suggests the presence of a dominant cyclical structure associated with the daily variation in solar irradiation. This behavior aligns with the expectation that irradiation is null at night, gradually increases throughout the day, reaches a peak around noon, and decreases until sunset. The approximately elliptical or circular shape observed, for example in the pairs 6×7, and 8×9 indicates a relationship between these components and a predominant faster cycle.

As we move further into the matrix, some pairs display more complex patterns, with multi-sided polygonal shapes, pairs 12x13 and 20x21, for example. These patterns indicate the presence of variations that are not limited to the daily cycle but also include medium-term oscillations, possibly associated with intra-seasonal and monthly variations. This is consistent

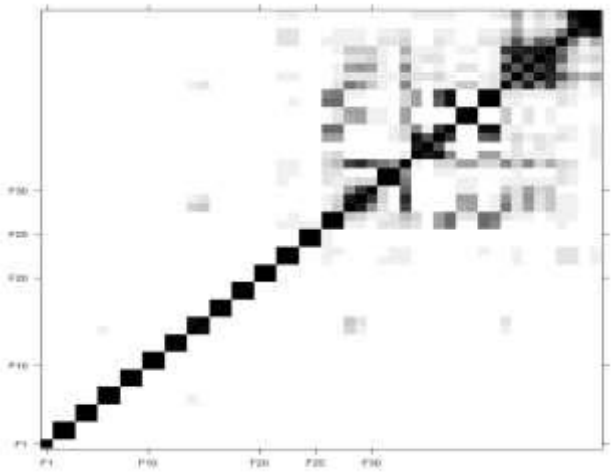
with the observation that solar irradiation variability tends to be higher during summer months, when there is greater radiation availability and increased atmospheric turbulence, and lower during winter months, when days are shorter and irradiation remains more stable throughout the day.

Additionally, some of the more advanced pairs of eigenvectors exhibit highly irregular patterns or nearly homogeneous blue-filled areas, for instance pairs 28x29 to 49x50. These patterns indicate the presence of noise components or high-frequency fluctuations, which may be related to daily atmospheric variability, such as cloud cover, short-term meteorological oscillations, and other sources of interference that do not follow a well-defined pattern.

Another relevant aspect is that the polygons formed in certain pairs of eigenvectors reflect the periodicity associated with different time scales. Polygons with fewer sides may be linked to lower-frequency seasonal oscillations, such as annual variations in irradiation due to seasonal changes. On the other hand, more complex and enclosed shapes may be associated with short-term phenomena, such as rapid meteorological changes.

Therefore, the decomposition of the time series through SSA enables the identification of multiple seasonal components in solar irradiation, ranging from daily variability to seasonal variations throughout the year. The analysis of the pairs of eigenvectors and their respective graphical representations suggests the presence of well-defined cyclical patterns, reinforcing the influence of astronomical and atmospheric factors on the dynamics of solar irradiation.

Figure 21 - W Correlation Matrix



Source: Author's own elaboration

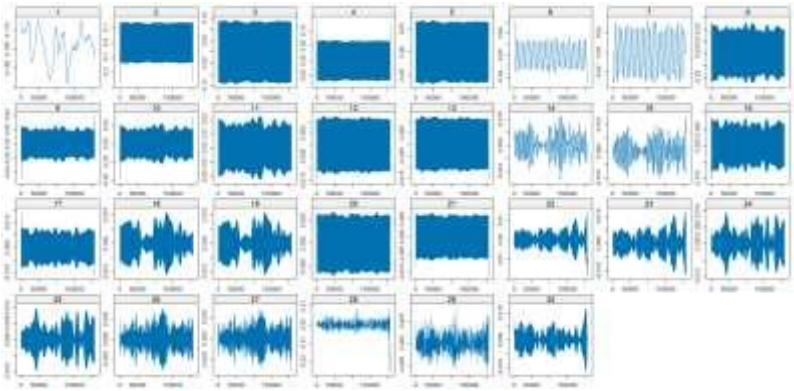
Figure 21 presents the weighted correlation matrix W, which aids in classifying the components to be used in the series reconstruction. This matrix displays correlations among the factors on a continuous scale, where darker shades indicate high correlations (close to 1), while lighter shades represent low correlations (close to 0). It can be observed that the first factor (F1) has a low correlation with the others, suggesting high separability from the remaining components. Furthermore, factors 2 through 20 show strong correlations not just among themselves, there is always a correlation between 2 and 3, 4 and 5, 6 and 7 and so on, more or less halfway. This corroborates the analysis of paired eigenvectors given that those with the same seasonalities form a polygon and those with different seasonalities do not form a polygon (e.g.: 3 and 4, 5 and 6, 7 and 8). Beyond this range, a progressive decline in

correlation is evident, with the subsequent factors becoming more correlated with each other, potentially indicating the presence of noise.

Figure 22 - Reconstructed series up 50 eigentriple

Source: Author's own elaboration

Figure 23 - Reconstructed series up 30 eigentriple



Source: Author's own elaboration

Figures 22 and 23 present the reconstructed series graphs obtained in the final stage of SSA application, offering further insight into which eigentriples should be used to reconstruct each component of the series. For instance, series 6 and 7 exhibit cyclical behavior around a mean, suggesting they belong to the signal component. In contrast, series 28 and 29, characterized by high frequency and large variance, can be classified as part of the noise component.

It is crucial to highlight that the selection of eigentriples for each component must integrate multiple analytical metrics. Identifying the eigentriple corresponding to the trend, for example, involves factors such as having the highest log-eigenvalue, the highest relative contribution, and a first factor with low correlation with the others. Similarly, the selection of eigentriples associated with noise is determined by the lowest log-eigenvalue, low relative contribution, absence of identifiable patterns in eigenvector pairs, high correlation among factors, and characteristics such as high frequency and large variance in the reconstructed series.

At the conclusion of the graphical analysis of the SSA estimation results, the following classification is established: the first eigentriple represents the trend component, eigentriples 2

through 22 correspond to the cyclic component, and the remaining eigentriples are classified as noise.

4.3.3 Forecast Model

This section presents and analyzes the results obtained from the forecasting model. Regarding the forecasting horizon, a value of h = 24 was chosen, representing a complete daily cycle of solar irradiation. The primary objective of selecting this horizon is to capture fluctuations throughout the day, allowing for a detailed analysis of intra-day variations. This aspect is crucial, as solar irradiation exhibits dynamic and non-stationary behavior influenced by meteorological and seasonal factors. By using a 24-hour period, the aim is to gain a better understanding of short-term patterns, which can be critical for applications that rely on accurate forecasts, such as solar energy management.

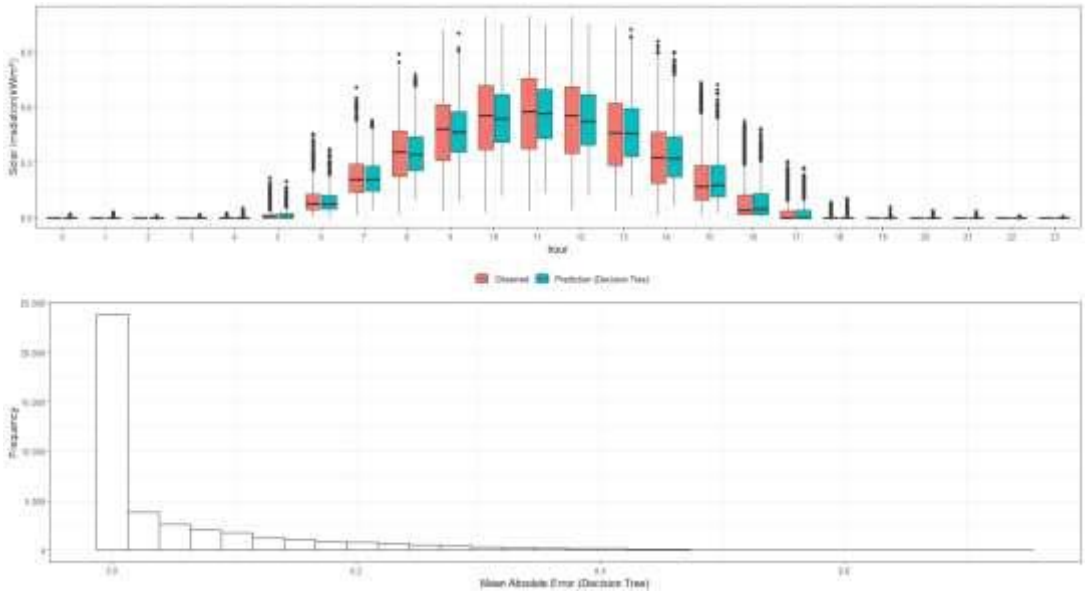
The results obtained for each model are summarized in Table 2. Firstly, it was observed that the benchmarking model performed worse compared to the other models. Analyzing the metric values, it is observed that, in both evaluations, the SSA + ANN model outperforms the SSA + Regression Tree model. This result suggests that combining the SSA decomposition technique with ANN enables the model to capture more relevant patterns in the time series, effectively reducing forecasting errors.

Table 3 - Error Metrics (Antonia(PR))

Model	RMSE	MAPE
ARIMA	2.670	0.717
SSA+ ANN	0,002	0,027
SSA + Regression Tree	0,014	0,066

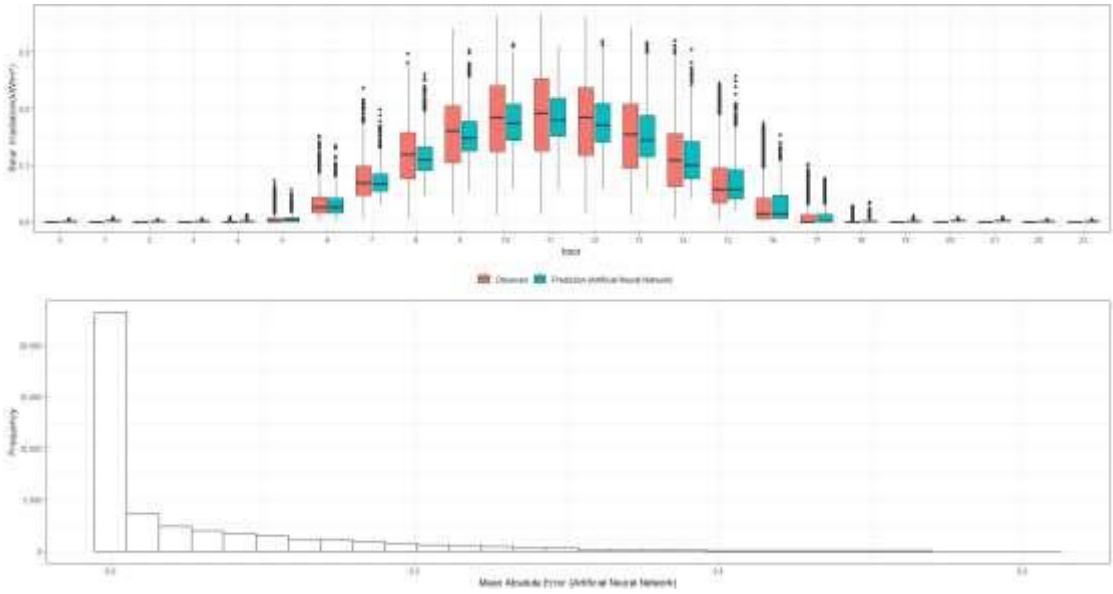
Source: Author's own elaboration

Figure 24 - Forecast Model Accuracy (SSA + MPL)



Source: Author's own elaboration

Figure 25 - Forecast Model Accuracy (SSA + Regression Tree)



Source: Author's own elaboration

Figures 24 and 25 illustrate the comparison between observed and predicted values obtained from the Regression Tree and Artificial Neural Network (ANN) models, respectively. In both cases, the first chart presents the distribution of observed (red) and predicted (blue) values throughout the day using boxplots. As expected, solar irradiation remains close to zero during nighttime hours (12 AM to 5 AM and after 6 PM), increases from 6 AM, reaches its peak between 10 AM and 2 PM, and gradually decreases after 3 PM.

Analyzing Regression Tree model, Figure 24, it is evident that the predictions closely follow the observed values, particularly during peak irradiation hours. The median of the predictions aligns well with the observed values, indicating that the model captures the general trend effectively. However, the observed data exhibit greater variation, and the

predicted values do not fully capture this fact, especially in the during peak hours periods. Additionally, some outliers are observed, especially during sunrise and sunset, which may be attributed to atmospheric conditions not captured by the model, such as cloud cover and sudden weather changes.

Similarly, in Figure 25, the Regression Tree model demonstrates strong predictive performance, closely following the observed values. However, compared to the ANN model, it exhibits slightly lower dispersion in transition periods, which may explain its slightly inferior performance in error metrics. The Regression Tree model appears to capture variations in irradiation well but still exhibits slightly larger errors in certain cases. As in the previous case, the observed data present greater variation, and the predicted values do not completely reflect this characteristic.

Overall, the ANN model demonstrates superior predictive performance, achieving lower RMSE and MAPE values while maintaining strong alignment between predicted and observed values.

5 Finals Remarks

This study developed indirect, short-term, hybrid hourly solar irradiation forecasting model. First it was evaluated the financial viability of Green Hydrogen (H_2) production across Brazilian municipalities using the Levelized Cost of Hydrogen (LCOH) metric of equilibrium price. The municipalities with the lowest (Xique-Xique, BA) and highest (Antonina, PR) costs were then selected. Subsequently, hourly time series data of solar irradiation from 2022 to 2023 was obtained for both municipalities. With this data in hand, the Singular Spectrum Analysis (SSA) technique was applied to decompose the series into trend, cyclical, and error components. Finally, the processed series was used in forecasting models.

The contributions of this work are significant. First, it offers a refined approach to assessing LCOH, incorporating precise solar irradiation forecasts to minimize uncertainty. Second, the study highlights regional disparities in LCOH across Brazil, emphasizing the role of site selection and renewable resource availability in determining economic feasibility. Finally, by demonstrating the accurate performance of the SSA + ANN model over traditional forecasting techniques, the study provides a practical framework for improving decision-making in the hydrogen sector.

The primary results underscore the relationship between LCOH variation and solar irradiation accuracy. Sensitivity analysis revealed that a 1% increase in average solar irradiation leads to a 0.437% reduction in LCOH, highlighting the direct impact of energy availability on hydrogen production costs. Additionally, the study identified substantial regional differences, with Xique-Xique (BA) exhibiting the lowest LCOH due to its favorable solar conditions, while Antonina (PR) displayed the highest costs. These findings reinforce the necessity of precise forecasting models, as even small inaccuracies in irradiation predictions can lead to suboptimal investment decisions, affecting financial planning and operational efficiency.

Compared to existing literature, this study advances the field by integrating sophisticated time-series forecasting techniques into LCOH assessments, bridging a crucial gap in understanding the economic viability of Green Hydrogen production in Brazil. While prior research has focused on global hydrogen production trends (Farrel,2023; Saker *et al*, 2023) or broader economic evaluations (De Andrade *et al, 2024*), this work provides a granular analysis at the municipal level, offering actionable insights for policymakers and investors.

Despite its contributions, this study has some limitations. The analysis primarily considers solar irradiation as the key determinant of LCOH, without accounting for other external factors such as grid infrastructure, fluctuations in water availability, and proximity to consumption centers or ports. Furthermore, in the forecasting model, only the variable solar irradiation was considered, while additional meteorological variables such as precipitation levels, temperature, air pressure, and cloud cover could enhance model accuracy and improve forecast reliability (Liu *et al*, 2022; Radzi *et al*, 2023).

In conclusion, this study provides a robust framework for evaluating Green Hydrogen production viability in Brazil, emphasizing the critical role of accurate solar irradiation

forecasting in economic assessments. The findings offer valuable insights for players seeking to optimize hydrogen production strategies, paving the way for a more sustainable and economically viable hydrogen economy.

REFERENCIES

ABSOLAR. Infográficos. Disponível em: https://www.absolar.org.br/mercado/infografico/. Acesso em: 5 de abril de 2024.

ANEEL. Resolução Normativa ANEEL Nº 482 DE 17/04/2012. Disponível em: https://www.legisweb.com.br/legislacao/?id=342518. Acesso em: 13 de abril de 2024.

ANEEL. Resolução Normativa ANEEL Nº 687 DE 24/11/2015. Disponível em: https://www2.aneel.gov.br/cedoc/ren2015687.pdf. Acesso em: 6 de setembro de 2024.

Aparna, S. (2018, June). Long short term memory and rolling window technique for modeling power demand prediction. In 2018 Second international conference on intelligent computing and control systems (ICICCS) (pp. 1675-1678). IEEE.

Asmelash, E., Prakash, G., Gorini, R., & Gielen, D. (2020). Role of IRENA for global transition to 100% renewable energy. *Accelerating the transition to a 100% renewable energy era*, 51-71.

Brown, D. P., Cajueiro, D. O., Eckert, A., & Silveira, D. (2024). Evaluating the Role of Information Disclosure on Bidding Behavior in Wholesale Electricity Markets. University of Alberta, Faculty of Arts, Department of Economics.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Brazil. *Law No. 10,438/2002, 2025*. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/l10438.htm. Acesso em: 20 de janeiro de 2025.

Cardoso, D. S., Locatelli, P. S., Ramalho, W., & Asgary, N. (2021). Distributed generation of photovoltaic solar energy: impacts of ANEEL's new regulation proposal on investment attractiveness. *Revista de Administração da UFSM, 14*, 423-442.

Ciaburro, G., & Venkateswaran, B. (2017). *Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles*. Packt Publishing Ltd.

De Andrade, J. V. B., da Costa, V. B. F., Bonatto, B. D., Áquila, G., de Oliveira Pamplona, E., & Bhandari, R. (2024). Perspective under uncertainty and risk in green hydrogen

investments: A stochastic approach using Monte Carlo simulation. *International Journal of Hydrogen Energy, 49*, 385-404.

Du, L., Lin, L., Yang, Y., Li, J., Xu, S., Zhang, Y., & Zhou, L. (2024). Environmental and economic tradeoffs of green hydrogen production via water electrolysis with a focus on carbon mitigation: A provincial level study in China. *International Journal of Hydrogen Energy*.

Farrell, N. (2023). Policy design for green hydrogen. Renewable and Sustainable Energy Reviews, 178, 113216.

Governo. Matriz Elétrica Brasileira alcança 200GW. Disponível em: https://www.gov.br/aneel/pt-br/assuntos/noticias/2024/matriz-eletrica-brasileira-alcanca-200-gw. Acesso em: 5 de abril de 2024.

Golyandina, N., Korobeynikov, A., & Zhigljavsky, A. (2018). *Singular spectrum analysis with R*. Springer Berlin Heidelberg.

Gruszczynska, M., Klos, A., Rosat, S., & Bogusz, J. (2017). Deriving common seasonal signals in GPS position time series: By using multichannel singular spectrum analysis. *Acta Geodynamica et Geomaterialia, 14*(3), 267-278.

Hassani, H., & Zhigljavsky, A. (2009). Singular spectrum analysis: methodology and application to economics data. *Journal of Systems Science and Complexity, 22*, 372-394.

Hassani, H., Mahmoudvand, R., & Zokaei, M. (2011). Separability and window length in singular spectrum analysis. *Comptes rendus mathematique, 349*(17-18), 987-990.

Hassani, H., Soofi, A. S., & Zhigljavsky, A. (2013). Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society, 176*(3), 743-760.

Hassani, H., Webster, A., Silva, E. S., & Heravi, S. (2015). Forecasting US tourist arrivals using optimal singular spectrum analysis. *Tourism Management, 46*, 322-335.

Hochreiter, S., et al. (1997). Long Short-term Memory. *Neural Computation*. MIT Press.
Hua, X., et al. (2015). Short-term power forecasting for photovoltaic generation based on wavelet neural network and residual correction of Markov chain. *IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 1-5.

IEA. (2023). *Global Hydrogen Review 2023*. Disponível em: https://iea.blob.core.windows.net/assets/ecdfc3bb-d212-4a4c-9ff7-6ce5b1e19cef/GlobalHydrogenReview2023.pdf. Acesso em: 20 de janeiro de 2025.

IRENA. (2019). *Hydrogen: A Renewable Energy Perspective*. Disponível em: https://www.irena.org/publications/2019/Sep/Hydrogen-A-renewable-energy-perspective. Acesso em: 15 de janeiro de 2025.

Lee, H., Calvin, K., Dasgupta, D., et al. (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report*. IPCC, Geneva, Switzerland.

Li, Z., et al. (2016). A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting. *Energies, 9*(1), 55.

Liu, C., et al. (2022). A review of multitemporal and multispatial scales photovoltaic forecasting methods. *IEEE Access, 10*, 35073-35093.

Lopes, E., et al. (2023). Desenvolvimento de um Roadmap técnico-comercial da energia solar fotovoltaica centralizada no Brasil. *11th Seminar on Power Electronics and Control (SEPOC 2018)*.

Maestre, V. M., Ortiz, A., & Ortiz, I. (2021). Challenges and prospects of renewable hydrogen-based strategies for full decarbonization of stationary power applications. *Renewable and Sustainable Energy Reviews, 152*, 111628.

Mohamad Radzi, P. N. L., Akhter, M. N., Mekhilef, S., & Mohamed Shah, N. (2023). Review on the application of photovoltaic forecasting using machine learning for very short-to long-term forecasting. Sustainability, 15(4), 2942.

Oliveira, D. S. (2022). *Previsão horária para o consumo de energia elétrica no Brasil considerando a contribuição da geração distribuída fotovoltaica* (Doctoral dissertation, PUC-Rio).

ONS. *O que é SIN*. Disponível em: http://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin. Acesso em: 5 de abril de 2024.

Parmezan, A. R. S., Souza, V. M. A., & Batista, G. E. A. P. A. (2019). Evaluation of

statistical and machine learning models for time series prediction. *Information Sciences, 484*, 302-337.

Pereira, E. B., et al. (2017). *Atlas brasileiro de energia solar*. São José dos Campos: Inpe.

Rady, E. H. A., Fawzy, H., & Fattah, A. M. A. (2021). Time series forecasting using tree-based methods. *J. Stat. Appl. Probab, 10*(1), 229-244.

Santos, M. F. D. O. (2022). *Singular spectrum analysis para estimativa de núcleo de inflação no Brasil*.

Santos, M. T. (2022). *Estudo sobre a geração distribuída no Brasil: benefícios e impactos para distribuidoras e consumidores*.

Sarker, A. K., et al. (2023). Prospect of green hydrogen generation from hybrid renewable energy sources: A review. *Energies, 16*(3), 1556.

Tavares, L. A. (2023). *Matriz elétrica brasileira e as tendências futuras*. *RECIMA21*, 4(5), e453135.

Villanueva, B., et al. (2023). Cooperativas de energia no Brasil: um referencial teórico para a transição energética. *DELOS: Desenvolvimento Local Sostenível, 16*(50), 4061-4083.

Xiao, Z., et al. (2023). A novel method based on time series ensemble model for hourly photovoltaic power prediction. *Energy, 276*, 127542.